

INTONATION SYSTEMS
 IN
SCOTTISH ENGLISH

Karen Lois Currie

Ph.D.
University of Edinburgh
1979

To my mum and dad
Gail and John Currie

Acknowledgements:

I would like to thank the S.S.R.C. for their financial support and for allowing me to use the data collected for their project HR3601/1 entitled The Intonation of Scottish English. I would also like to thank my supervisor Dr. Gillian Brown for her continual encouragement and stimulation. I also appreciated her criticism of the countless drafts which passed through her hands. I would like to thank George Yule for the time and energy which he expended in discussions about the theoretical issues and implications involved in this thesis. I would also like to thank previous colleagues who worked with me in the project -- Joanne Kenworthy, Neil Mitchison and Roger S. Brown.

A data-based thesis cannot be written without those people who supply the data. Many thanks must go to my informants, specifically Margaret Richards, Ron Motherwell, and the phoneticians and students at Edinburgh University who suffered the agonies of the 'tonic' experiments.

A thesis must not only be written but it must also be submitted in a presentable form. I would like to thank John H. Currie, who laboured for many hours to produce the majority of the figures in this thesis, and also Gail Currie for providing 'calm' at moments of crisis.

Lastly, I would like to thank my husband, Mauricio Brito de Carvalho who not only typed this thesis, but who cooked for me, cleaned for me, washed for me and above all cared for me --- special thanks to Mauricio.

Table of contents:

	page
Chapter 1.	
1.0 Introduction	1
1.1 Acoustic correlates	2
1.2 Physiological correlates	8
1.2.1 Sub-glottal air pressure	8
1.2.2 Muscular activity -- electromyography....	11
1.2.3 Muscular activity -- laryngograph	14
1.3 The usefulness of physical correlates ...	16
Chapter 2.	
2.1 Duration perceived as length	20
2.2 Intensity perceived as loudness	23
2.3 Fundamental frequency as pitch	27
2.4 Psychophysical and linguistic judgments .	31
2.4.1 Linguistic judgments of a semantico- syntactic nature	32
2.4.2 Judgments of an attitudinal nature	42
Chapter 3.	
3.1 Halliday's system of intonation	54
3.2 Crystal's system of intonation	74
Chapter 4.	
4.1.0 Data collection	96
4.1.1 Text	97

4.1.2	Sentences read aloud	99
4.1.3	Questionnaire	102
4.1.4	Echo-question elicitation	104
4.1.5	Spontaneous speech	106
4.1.6	Game situation	107
4.2	Analysis of the data	108
4.3	Initial examination of the data	111
4.4	Pause phenomena in the data	119
4.5	Planning phenomena	135
4.6	Pitch movement in the data	145
Chapter 5.		
5.0	Introduction	153
5.1	Experiment 1: hypothesis	159
5.1.1	Method	159
5.1.2	Sentences containing the cumulation (ABC)	166
5.1.3	Sentences where the PM are spread over several lexical items	177
5.1.4	Sentences where the rightmost lexical item combines with phonetic maxima ...	181
5.1.5	Conclusion	183
5.1.6	Judge groupings	184
5.1.7	Overall results	188
5.2	Experiment 2	189
5.2.1	Method	193
5.2.2	Results: sentences containing the	

	cumulation (ABC)	195
5.2.3	Conclusion	204
5.3	Experiments 3, 4, 5	205
5.3.0	Method	205
5.3.1	Hypothesis	207
5.3.2	Results: experiment 3	207
5.3.3	Results: experiments 3,4,5	219
Chapter 6.		
6.0	An initial examination of contour units in readings of a text	258
6.1	Pause duration measurement	265
6.2	Fundamental frequency characteristics .	271
Chapter 7.		
7.0	An initial examination of contour units in spontaneous speech	359
7.1	Contour units in spontaneous speech: Edinburgh English	362
7.1.1	Pause measurement	362
7.1.2	F ₀ measurement	367
7.1.3	Summary of findings in Edinburgh spont- aneous speech	380
7.2	Contour units in spontaneous speech: Thurso English	382
7.3	Contour units in spontaneous speech: Glasgow English	385
7.4	Conclusion	394

Chapter 8.

8.0	Introduction	402
8.1	The tonetic stress marks system	403
8.2	Pitch levels system	405
8.3	The proposed model for Scottish English	409
8.3.1	The fundamental system of intonation .	411
8.3.2	The tone unit system	415
8.3.3	The tone group system	415
8.3.4	The sense group system	417
8.3.5	The paratone system	418
8.4	The realisation of the model	419
8.5	Support for contour system	433
8.6	Contour system in other languages	435

Abstract

This thesis is a data-based examination of the intonation system in Scottish English. As an introduction to the thesis I examine briefly the physical and psychophysical correlates of intonation. I then go on to discuss the work of two of the main investigators in the field of British intonation -- D. Crystal and M.A.K. Halliday. An initial analysis of the data-base with reference to the analyses proposed by Crystal and Halliday led me to set up a series of experiments to test the reality of the notion 'tonic'. These experiments are described in detail.

Six readings of a text are then examined with specific reference to pause duration and fundamental frequency measurements. The results of the 'tonic' experiments and the measurements of the texts prompted me to propose a contour system analysis of intonation. I propose that there is a neutral contour in Scottish English (specifically Edinburgh Scottish English) which is typified by two stressed peaks of prominence which deviate from a baseline of unstressed syllables. This contour varies its realisation according to its function eg. the initial peak is boosted when a new topic is introduced.

Support for such a contour system analysis is provided from the literature of various languages (specifically Dutch, German, Russian and Danish).

Chapter 1.

1.0 Introduction

This thesis is concerned with a study of the systems of intonation realised in Scottish English. Before we examine the system of intonation, however, it is necessary to examine the characteristics of intonation itself. This chapter will deal briefly with the physical correlates of intonation and the following chapter will deal with the perception of intonation, i.e. the psychophysical correlates of intonation.

There are two main schools of thought involved in current investigation of the physical correlates of intonation, and each school is oriented differently. The main school consists of people such as Fry, Lehiste, Denes, Milton-Williams, etc. who investigate the acoustic wave form of the speech signal and correlate intonation with features such as intensity, duration, amplitude and fundamental frequency. Those people who investigate the acoustics of speech might be said to be Hearer oriented.

An alternative viewpoint is to investigate the physiological correlates of intonation such as laryngeal vibration and sub-glottal air pressure. Investigators in this field might be said to be Speaker oriented (e.g. Lieberman, Liberman, Cooper, van der Berg, Fourcin, etc.).

1.1 Acoustic correlates

The sound spectrograph is the most frequently used technique for investigating the acoustic correlates of the speech signal. The sound spectrograph is a wave analyser which produces a permanent visual record showing the distribution of acoustic energy in both frequency and time. The frequency scale of the spectrogram is linear and covers 3,500 cycles. The time scale is also linear and covers 2.4 sec. Filters can be used to examine a specific range of frequencies and the time scale can also be altered by recording at faster or slower speeds.

It is generally agreed that intonation is realised acoustically by the variation in fundamental frequency. But when theories of intonation are established, the notion of stress also plays an important part -- both in the American theories of pitch accent, and in the British theories of nuclear tone movement description. Both of these types of theory involve the notion of stress as well as the notion of variation in pitch movement.

There are various conditioning factors which affect the fundamental frequency at which a syllable nucleus is realised. These factors include intrinsic pitch, preceding and following sounds, and other suprasegmental features, especially stress. The term intrinsic pitch refers to the fundamental frequency which is determined by the phonetic quality of a vowel, i.e. there is a connection between vowel quality and the relative height of

the average fundamental frequency associated with it. In an investigation carried out by Lehiste and Peterson (1961) the following average fundamental frequencies were found in a total of 1263 syllable nuclei examined. The nuclei all occurred at the peak of an intonation contour to keep the context constant, and all were produced by a single speaker in order to avoid inter-speaker variation (since the fundamental frequency varies from person to person). The average F_0 for each nucleus studied is given below:

/i/	183Hz	/u/	182Hz
/I/	173Hz	/U/	171Hz
/e/	169Hz	/o/	170Hz
/ε/	166Hz	/ɔ/	165Hz
/ae/	162Hz	/a/	163Hz

Similar results have been obtained for various other languages, for example, Serbo-Croatian and Itsekiri, a West African tone language (cf. Lehiste, 1961).

Preceding consonants have very little influence on the duration of the following vowels, but they certainly influence the fundamental frequency of the following syllable nuclei. Higher fundamental frequencies are associated with voiceless consonants, so much so, that the influence of an initial consonant may counter-balance the influence of the intrinsic pitch. In the

study carried out by Lehiste and Peterson mentioned above (1961), it was found that the average for /kae/ sequences was 171 Hz, while that of /gi/ sequences amounted to 170 Hz. The difference in average peak values due to the voicelessness or voicing of the preceding consonant is often accompanied by a different distribution of the F_0 movement over the word such that after a voiceless consonant, and especially after a voiceless fricative, the highest peak occurs immediately after the consonant. After a voiced consonant, on the other hand, the fundamental frequency tends to rise slowly and the peak will occur approximately in the middle of the test syllable (see Lehiste and Peterson 1961).

Final consonants do not seem to have any systematic influence on the fundamental frequency of the syllable nucleus preceding it except that the tonal movement may continue during a following voiced resonant. Thus we may regard sequences of vowels and post-vocalic resonants as complex syllable nuclei whose F_0 is distributed over the whole sequence, in a similar way to the tonal distribution over diphthongs.

Fundamental frequency is also affected by stress, such that a stressed syllable is frequently associated with a higher fundamental frequency.

Stress has been widely investigated and its acoustic correlates have formed the bases for countless

experiments. One of the leading investigators in the acoustic field is D. B. Fry, and in 1958 he described three experiments which he carried out using the acoustic parameters of duration, intensity, and fundamental frequency. The tests were carried out on word pairs of the type subject, object, and digest. (It should be noted here that the majority of experiments involving stress have been carried out using single words). In the first experiment variations in duration were combined with variation in intensity and the results of this experiment showed that both duration and intensity acted as cues in stress judgments. The second experiment combined variations in duration with step changes of fundamental frequency. This experiment showed that the change in frequency had a marked effect on the judgments but that the magnitude of the frequency change had no marked effect, i.e. if a syllable was high enough to be heard as stressed it would be heard as stressed from that threshold on. Further increases in F_0 value made no difference. The third experiment included variation in F_0 within one syllable and contained a range of patterns which imposed sentence intonation on the test items. Fry says about this last experiment:

"The results again demonstrated the all-or-none effect of frequency changes and showed that this may

outweigh the duration cue altogether." (1958, p.126)

Thus since fundamental frequency cues seem to be the main cues for recognising stressed items, perhaps the only acoustic correlate of intonation is fundamental frequency:

"in taking in an intonation pattern, he (the listener) may commonly rely mainly on variation in pitch." (Fry, 1958, p. 127)

Other experimenters have produced slightly different results. For example many people have emphasised the relationship between fundamental frequency and intensity (Denes 1959, Draper et al 1959, Ladefoged 1961, Denes and Milton-Williams 1959, 1962). This relationship has normally been agreed to be ^a complementary one, however, rather than each parameter functioning independently. A few investigators, nevertheless, have suggested that intensity (or loudness in the case of Crystal 1969) can operate as the main cue for stress. Denes and Milton-Williams found that this was a possibility:

"The conclusions that could be reached from these comparisons were that fundamental frequency, as well as intensity and duration, could be used by

listeners to make intonation judgments, but that fundamental frequency provided the dominant cue in those intonation groups associated with large frequency changes. For other categories of intonation, confusion resulted from the opposition of fundamental frequency and other cues, but when both intensity and duration were in agreement then the frequency cues were overpowered." (Denes and Milton-Williams, 1959, p. 11)

Therefore fundamental frequency seems to operate as the strongest cue to intonation judgments, normally operating with the 'support' of the other acoustic cues, but when these cues are in opposition (normally achieved only in synthetic speech situations) then the F_0 cue may be 'overpowered'. Denes and Milton-Williams have this to say in conclusion:

"The transmission and perception of the information associated with intonation is carried on the acoustic level by a complex pattern of fundamental frequency, intensity and duration cues. It also seems probable that the information carried by intonation is not encoded solely in acoustic dimensions of fundamental frequency, intensity and duration, but that sentence structure and context, etc., must also play their part." (1959, p. 13-14)

1.2 Physiological correlates

1.2.1 Sub-glottal air pressure

One of the most prominent investigators in the field of sub-glottal pressure is P. Lieberman, and throughout this section I will draw largely from his book Intonation, Perception and Language (1967). Lieberman believes that speech is organised in terms of expiratory airflow from the lungs involving the coordinated activity of several groups of muscles in the chest and abdomen. At the end of each expiration the flow of air out of the lungs ceases, and the subglottal air pressure abruptly falls. Lieberman states:

"the fundamental frequency of phonation is directly proportional to the subglottal air pressure. The other parameters that can affect the fundamental frequency of phonation are the tension and the phonation neutral position of the vocal cords. If the tension of the laryngeal muscles remains constant, then the fundamental frequency of phonation will fall at the end of the expiration." (1967, p. 26)

Lieberman notes that it is a universal of human speech that the fundamental frequency of phonation and the acoustic amplitude fall at the end of a sentence except in certain predictable cases (see chapter 6 for confirmation of this point but chapter 7 for evidence

against this conclusion from Glasgow speech). He suggests that the physiological basis of this phenomenon may be a condition of least articulatory control, i.e. if the tension of the laryngeal muscles is not deliberately increased at the end of expiration when the subglottal air pressure falls, the fundamental frequency will also fall. Lieberman says p. 27:

"This pattern of articulatory activity thus produces a prosodic pattern that is characteristic of the ones that are used to delimit the boundaries of unemphatic, declarative sentences in normal speech. We shall term this pattern of articulatory activity the 'archetypal normal breath-group'."

Lieberman suggests that breath groups of individual languages might be characterised by different patterns of laryngeal tension control in the non-terminal parts of the breath group. In British English, for example, the tension of the laryngeal muscles may be high initially and gradually fall throughout the breath group, whereas in American English the laryngeal tension remains relatively constant throughout the breath group.

A marked breath group is also argued for. The marked breath group contrasts with the unmarked breath group during the last 150-200 m/sec of phonation where

the tension of the laryngeal muscles increases in the marked breath group. The increased tension of the laryngeal muscles counters the falling subglottal air pressure, and the marked breath group thus has a terminal not-falling fundamental frequency contour. The marked breath group therefore can be regarded as the 'simplest' alternative to the unmarked breath group since the laryngeal tension is only increased at one point in the breath group, i.e. when the subglottal air pressure falls. Thus the terminal rise can be said to be the 'marked' alternative to the unmarked fall.

It should be noted at this point that speakers can use other articulatory manoeuvres in order to produce breath groups that are acoustically or perceptually similar to the prosodic patterns that result from the archetypical articulatory manoeuvres. For example speakers can produce a breath group where the subglottal air pressure function is similar to the archetypical patterns without the speaker pausing for inspiration at the end of the breath group. Lieberman does not specify what these other articulatory manoeuvres might consist of however.

1.2.2 Muscular activity - electromyography

The above discussion has been concerned with measurements of the subglottal air pressure which results from the coordinated activity of several groups of muscles. The activity of these muscles can also be measured, for example by means of electromyography. A full account of this technique is given in Buchtal (1957) and also in Ladefoged (1967). Briefly, all the muscles which are under voluntary control consist of a large number of elongated cells or fibres, each of which can contract. These fibres are organised into groups, each group being connected to a specific nerve cell in the spinal cord. This combination of a nerve cell and its associated muscle fibres is known as a motor unit. Each of the muscles involved in respiration for example consists of many thousands of motor units. When the muscle fibres which make up a motor unit are stimulated, an electric discharge known as an action potential is produced by the muscle cells. It is possible to record this electrical activity in the muscle in two ways. In the first method, electrodes consisting of silver plates about 5mm in diameter are placed on the skin immediately above the muscle; in the second method a hollow needle containing an insulated central wire is inserted into the muscle. In either case the electric potential occurring between the surface electrodes or between the insulated wire and the outer shaft of the needle can be

amplified and displayed on a cathode ray oscilloscope thus providing permanent records of the muscular activity.

Ladefoged (1958, 1967) investigated the activity of various muscles by means of electromyography, and as a result found that the respiratory muscles which are most active during speech are the internal intercostals.

The burst of activity of the internal intercostals were first noted by Stetson (1951) who concluded that

- i) Every syllable is accompanied by a 'ballistic chest pulse' produced by the action of the internal intercostal muscles.
- ii) In a stressed syllable the action of the intercostal muscles is reinforced by the abdominal muscles, led by rectus abdominis.

Ladefoged disagreed with the above findings noting that there was an increase in the degree of muscular activity immediately before syllables which were heard as strongly stressed, but that each segment of speech which is perceived as a syllable is not necessarily accompanied by a separate burst of muscular activity. Ladefoged notes that not only can the tension of the intercostal muscles be varied over a large range, but also there can be variations in the rate of change of tension. Thus a single increase in tension can span

a group of articulations including two vowels separated by a consonant closure, and sometimes two separate bursts of activity may occur during what would normally be regarded as a single syllable. Thus Stetson oversimplifies the situation. With regard to Stetson's second conclusion concerning the reinforcement of stressed syllables by activity in the abdominal muscles, Ladefoged notes that this does not in fact happen except perhaps in cases of very emphatic stressing when the pressure in the lungs may be unusually high. In normal conversational English Ladefoged says, the abdominal muscles are in action only at the end of a very long utterance. Ladefoged summarises such physiological correlates of stress as follows:

"It is generally agreed that the stress of an utterance cannot be correlated with any single acoustic property. Thus a syllable which is perceived as having a strong stress does not necessarily have a greater intensity, nor a higher nor lower frequency, nor a longer duration, nor a specific quality. But on the basis of our evidence it appears that the degree of stress is often related to the extra increase in muscular activity. Thus differences in stress, like many consonantal differences can be ordered more simply in terms of the human behaviour producing them than in terms of the accompanying

acoustic phenomena." (1973, p. 211)

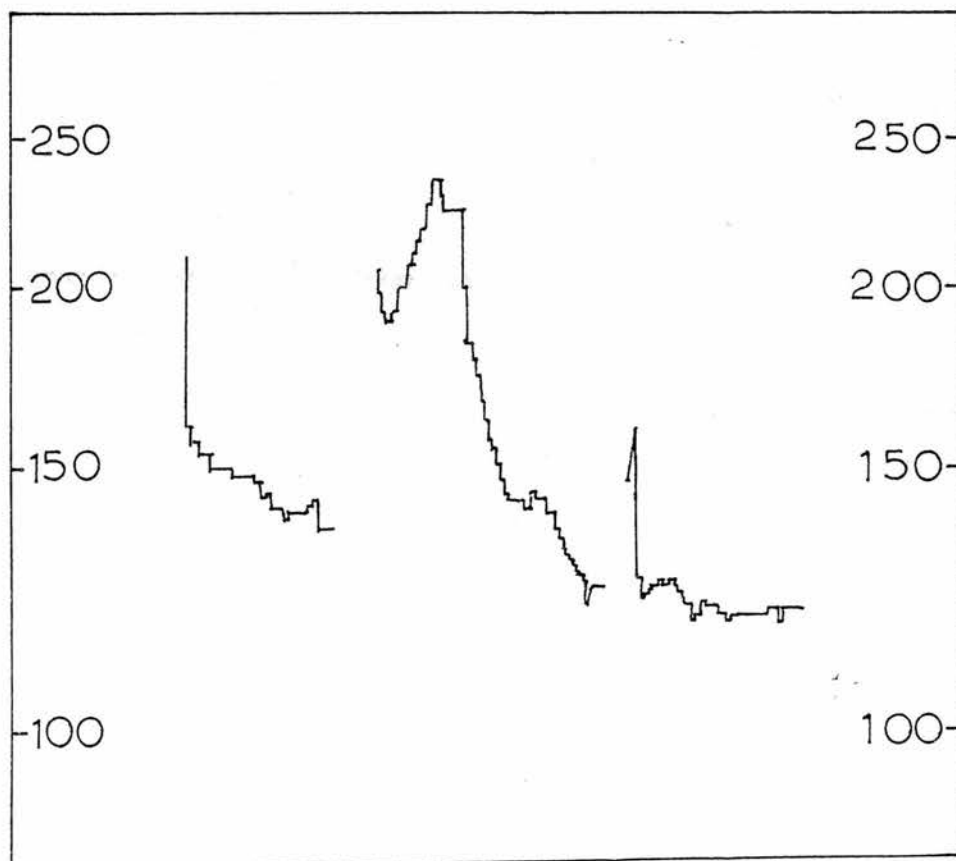
1.2.3 Muscular activity - Laryngograph

The technique of using electrical impedance to measure activity of some kind is also used to measure the activity of the vocal folds. The main investigator in this field is A. J. Fourcin, and in a paper entitled 'Laryngographic Examination of Vocal Fold Vibration' (1974) details of this technique are explained. The investigations which Fourcin describes in the paper involve the use of an electrical impedance technique for the direct examination of vocal fold closure, which does not interfere with phonation. This technique and its applications is described in many papers (cf. Abberton and Fourcin, 1971, 1975, 1976; and Simon, 1974). A brief summary of the methodology involved is given here.

Two guard-ring electrodes are applied superficially to the skin of the neck above the thyroid cartilage, and the laryngograph then acts as an impedance measuring device which measures the impedance across these two electrodes. The circuit has been designed so that the output from this electrode arrangement is self-compensating for speaker impedance variation (since the electrical condition of the neck can vary from person to person) and the instrument responds only to the fast changes associated with the vibration of the vocal

folds in ordinary vibration. The frequency rate of the vocal fold vibration (according to Fourcin) determines the pitch of the voice and there is a direct correlation between this frequency and intonation.

The laryngograph may be linked to an oscilloscope, and an oscilloscope trace will then display the fundamental frequency instantaneously on a logarithmic scale against time. This contour can then be plotted at either of two speeds (fast or slow) depending on the amount of detail required. This plotting device which is also linked to the laryngograph provides a permanent record of the contour. An example of the output is shown in Fig. 1.



This technique was used occasionally to supplement the auditory ^{analysis} in this thesis.

1.3 The usefulness of physical correlates

The school of acoustic investigators believe that the signals which they investigate form the basis of the input to the Hearer's system enabling the hearer to process speech. The school of investigators involved in examining the muscular activity and the resultant air pressures etc., believe that these physiological correlates are more realistic correlates upon which the hearer bases his analysis of speech (cf. Ladefoged, 1961). In his book called The Senses Considered as Perceptual Systems (1968) J.J. Gibson has this to say about the auditory system:

"A pure tone can be experienced only by neglecting transients and ordinarily only with an artificial stimulus. It can be produced with an audio-oscillator at a given frequency and a given amplitude, acting for a certain length of time at a fixed distance from an ear. The result is said to be a meaningless sensation having a pitch corresponding to the frequency, a loudness corresponding to the amplitude, and a certain duration. Meaningful sounds, however, vary in much more elaborate ways than merely in pitch, loudness and duration.

Instead of simple duration, they vary in abruptness of beginning and ending, in repetitiveness, in rate, in regularity of rate, or rhythm, and in other subtleties of sequence. Instead of simple pitch, they vary in timbre or tone quality, in combinations of tone quality, in vowel quality, in approximation to noise, in noise quality, and in changes of all these in time. Instead of simple loudness, they vary in the direction of change of loudness, the rate of change of loudness, and the rate of change of change of loudness. In meaningful sounds, these variables can be combined to yield higher-order variables of staggering complexity. But these mathematical complexities seem nevertheless to be the simplicities of auditory information, and it is just these variables that are distinguished naturally by an auditory system." (Gibson, 1968, p. 87)

Later in the same chapter Gibson discusses the relationship between the auditory stimulus and the muscular activity:

"any single elementary contraction does not correspond to any single elementary sensory datum. But the pattern of contractions and the change of nervous output at the muscles do correspond to the pattern of excitation and the change of nervous in-

put at the cochlea. There is no need to be associated, since they are identical ... what probably happens during the babbling stage is the differentiating of pattern and change in the muscular output along with a parallel differentiating of pattern and change in the cochlear input.

Another old formula has it that when an individual hears his own speech he gets the same stimuli as when he hears the speech of another. If this were literally true it would lead to incredible confusion -- the individual could not tell whether he or someone else had spoken! Actually the wave train invariants of stimulation are the same in the two cases, but the binaural balance of the stimulation is not; the symbols are perceived as the same, but the voices are perceived as different self-stimulation is binaurally balanced and unaltered by head-turning, whereas external stimulation is made to balance only by orienting the head." (Gibson, 1968, p. 95)

From the above lengthy quotations we can see that many more parameters are involved in what might be regarded as the physical correlates of intonation than might at first be supposed, and that all physical correlates must be related to perceptual correlates in order to abstract the patterns or symbols of speech

from the overall acoustic or muscular data, i.e. not everything that can be measured is relevant, and in order to differentiate between relevant and irrelevant information some sort of equivalence must be found between the physical signal and the perceptual symbol.

The perceptual correlates of stress and intonation will be examined briefly in the next chapter.

Chapter 2.

In this chapter we will examine the psychophysical correlates of some of the acoustic and physical parameters mentioned in the preceding chapter. The acoustic parameters mentioned were duration, amplitude or intensity, and fundamental frequency. The psychophysical correlates of these parameters are length, loudness, and pitch respectively. Let us concentrate our discussion on each of these parameters in turn.

2.1 Duration perceived as length.

It is very difficult to know how duration plays a part in intonation since the majority of experiments involving duration have been primarily concerned with the distinction between stressed and unstressed syllables. The duration of a stressed syllable has therefore been readily compared with the duration of an unstressed syllable. Such methodology is all very well when we are concerned with monosyllables or a direct comparison between two syllables but very little experimentation has been attempted involving syllable duration of meaningful forms in real speech. This is understandable since in order to compare the duration of one syllable with another all other parameters should ideally remain equal. In real speech parameters such as intrinsic vowel length, consonant length etc.

may affect judgments if we compare syllables with a different phonemic make-up from each other (cf. Lehiste and Peterson, 1961). Thus in order to determine whether a specific syllable is being perceived as 'longer' than usual, one would have to know what the normal length or duration of that syllable would be, and also whether the syllable could vary its duration according to context without necessarily being perceived as being different. One would therefore have to define a specific syllable as possessing a phonemic 'length' which could permit several allophonic variations of that length defined by context before one perceived the phoneme 'length' changing to '+ length' or '-length'.

Nooteboom and Slis (1969) carried out an experiment involving duration. They studied the effects of stress on a three syllable nonsense word which was produced consistently in the context of the same carrier sentence at all times. The three syllables were exactly the same phonemically which allowed the duration parameter to be treated as the single main variable. The nonsense words were of the form /pVpVpVp/ to allow the duration of both the vowel and the voiceless stop to be measured precisely without the confusion which might have resulted if a voiced consonant had been used. Nooteboom and Slis found that the duration of the stressed vowel was not influenced very much by changing the position of the stress within the word. The stres-

sed vowel seemed to have an optimal duration for the particular speech rate and word length under discussion. The unstressed vowels, however, showed considerable differences according to the position within the word. The vowel in the unstressed second syllable was considerably shorter than the vowel in the unstressed first syllable, whereas the vowel in the unstressed last syllable was much longer and in fact almost as long as the stressed vowel. These findings refer to the long vowel /a/. A short vowel was then substituted and different findings resulted such that in the first and final syllables a stressed short vowel did not differ significantly from an unstressed one. These findings indicated that the stressed short vowel showed in all positions a fixed difference from the stressed long vowel. Thus the short vowel cannot be easily lengthened under stress. The durational opposition between long and short vowels is optimised under conditions of stress. Nooteboom and Sils then go on to show that the observed long duration of the unstressed last vowel if it is present seems to be a characteristic of the rhythmic figure of the Dutch word. If it is the case that an unstressed last vowel is predictable according to the rhythmic structure of Dutch, then presumably length cannot act as a cue for stress in such a position.

A much more detailed investigation of syllable duration is necessary therefore in order to establish

rhythmic constraints within a single language and across various languages. The inherent duration characteristics of phonemes and their allophones, the inherent duration characteristics of phonemes in combination and isolation, and other such 'constants' must be established before we can attempt to establish 'length' variables which might function as cues for stress recognition and intonation.

2.2 Intensity perceived as loudness

Loudness is regularly believed to be the psychological correlate of sound intensity. Gulick (1971) notes that while it is obvious that under usual circumstances loudness grows with intensity, there are exceptions enough to require us to view loudness as a complex function of frequency and wave complexity as well as intensity.

Loudness can be measured by the discriminatory response of listeners, thus enabling a psychological scale of loudness to be set up whereby the numerical relationships would correspond to the sensation magnitudes. Among the psychophysical methods used to generate a loudness scale (and a pitch scale) the method of fractionation is prominent. This method uses a fixed reference tone (or in the case of setting up a pitch scale, a fixed intensity) which is then designated as having X units of loudness, the intensity is then

determined for a tone which is perceived to be half as loud, then this latter tone is halved and so on until a loudness scale can be determined with intensity correlates. Work on loudness scaling has been carried out by Churcher 1935, Churcher, King and Davies 1934, Fletcher 1934, Fletcher and Munsen 1933, Stevens 1935, 1936, Stevens and Davies 1938 and others.

The data available indicates that the lower the frequency, the more rapidly loudness increases with intensity, and since the scales suggested by Stevens and Davies and Churcher agree fairly closely, perceived scales of loudness would seem to be clearly definable as a function of intensity. Gulick (1971), however, points out that there are other factors which affect the perception of loudness such as frequency and complexity. The influence of frequency on loudness can be seen when two tones of different frequency are matched for loudness and their respective intensities are compared. There seems to be a general relationship such that as frequencies rise or fall from the middle region, their intensities must be progressively increased to maintain equality of loudness (cf. Fletcher and Munsen, 1933, and Steinberg and Munsen, 1936). Thus frequency influences the perception of loudness. The complexity of a tone will also affect the perceived loudness of that tone. Complex tones are perceived as louder than pure tones of equal sound pressure (cf. Fletcher 1934 and Pollack

1948) such that the growth in loudness with intensity is much more rapid for complex sounds than for pure tones. Gulick notes that the loudness of a complex tone depends, in part, upon the spectral distribution of its components, cf.:

"When the components are widely separated, the loudness of the tone is equal to the sum of the loudness of the components heard separately. However, when the components are close in frequency, the loudness is less than the sum of the individual loudnesses but still greater than that of a pure tone of the same sound pressure. The important point is that the loudness of a complex sound is not always the sum of the components. Failure to find simple additivity has led to speculation that loudness may depend somehow on place along the basilar membrane." (Gulick, 1971, p. 149)

Thus loudness cannot be said to be simply a function of intensity. Loudness is interrelated with frequency in a complex way, and more will be said about this relationship in the next section.

It should also be noted at this point that there is a difference between the perception of loudness from an 'external' input and from an 'internal' input. Lane, Catania and Stevens (1961) carried out experiments to

find out how an individual judged his own vocal production to see whether the primary cue was his perception of loudness in the form of a function of intensity, or whether the individual would use the cues of muscular effort. They say:

"It is of particular interest to compare the subjective scale for autophonic level with the subjective scale for loudness. The psychophysical law governing both modalities, speech and hearing, is a power function. Loudness, however, grows approximately as the 0.6 power of the sound pressure (S.S. Stevens, JASA 27, 185, 1955) whereas the autophonic scale grows approximately as the 1.1 power of the sound pressure. These two exponents tell us that, when the speaker raises his voice by what he judges to be a factor of 2, his voice will not sound twice as loud to a listener. In other words, there is a great difference from the subject's point of view between the relative subjective magnitude of sounds that he generates by his own vocal effort and those that are generated by an external source." (Lane, Catania and Stevens, 1961, p. 162)

From the above remarks, we see that not only is loudness a very complex psychophysical parameter involving frequency, intensity and wave complexity, but that

loudness is perceived differently according to the source. To state that loudness is the psychophysical correlate of intensity, then is a gross oversimplification of the facts.

2.3 Fundamental frequency as pitch

It has been amply demonstrated in the literature that pitch and frequency are not manifestations of the same phenomena (cf. Stevens and Davis, 1938; and Gulick, 1971). Gulick notes that these two terms, pitch and frequency, are frequently confused, and in some disciplines such as physics, the two terms are used interchangeably. However, it is necessary to separate the two here. Pitch is related to auditory sensation, while frequency describes one of the physical properties of sound. A tone with a specific frequency will only produce the sensation of tone if the frequency is heard, and that will depend on the sensitivity of the ear and the intensity of the tone. An everyday example of this phenomenon is the dog whistle which is not 'heard' by the human ear but which is 'heard' by the canine ear. This example demonstrates that the canine ear is more sensitive to higher frequencies than the human ear. The upper threshold of the human ear is lower than that of the canine ear and it is therefore unable to perceive frequencies which are perceived by the canine ear. Another point which

should be made is that two pure tones of different frequencies will be perceived as being identical in pitch if the differential sensitivity of the auditory system is insufficient to allow their discrimination. Pitch can also vary as a function of intensity even when frequency is constant. Thus, although pitch varies primarily as a function of frequency, we must differentiate between the two for the above reasons.

Psychophysical studies have demonstrated that pitch scales are related to frequency scales in a very complex fashion. Volkman and Stevens set up a pitch scale in 1937, then revised it in 1940. We will concern ourselves principally with the revised version. A pitch scale is obtained by assigning numeral to tones in such a way that the numerals stand in some definite relation to the pitch exhibited by the tones. Stevens and Volkman (1940) erected a subjective scale for the measurement of pitch. This scale was an extensive one, i.e. a scale whose numerals in the numeral series are reflected in perceived relations among the pitches. The unit of this scale is called a mel and it is defined as one thousandth of the pitch of a 1000 cps tone. This pitch scale was used to predict the differential sensitivity of the ear as a function of frequency. From the analysis as carried out by Stevens and Volkman (1940) they were able to conclude that all jnd's ('just noticeable differences') for pitch are es-

sentially equal in subjective size. It should be noted, however, that the jnd's for pitch are subjectively equal at a constant loudness level. On the other hand, an analogous treatment of similar data for the discrimination of intensity (Stevens and Davis, 1938) shows that jnd's for loudness are not subjectively equal. Instead the jnd's for loudness increase rapidly in size as the intensity of the stimulus is raised (cf. section 2.2). Stevens and Volkmann point out the difference between the two mechanisms as follows:

"... the physiological mechanisms which mediate our discriminations of pitch and loudness are of two basically different types. When the frequency of a tone is raised the pattern of excitation is shifted along the basilar membrane -- new excitation is substituted for old. Thus we tell one pitch from another because the locus of excitation on the basilar membrane has been altered. When we increase the intensity of a tone, however, the general locus of excitation in the cochlea remains relatively fixed and new excitation is added to old. Thus we detect that one tone is louder than another tone of the same frequency whenever there is an addition of sufficient excitation. Perhaps the subjective size of jnd's is related to the kind of discriminatory mechanism which mediates them." (Stevens and Volk-

mann, 1940, p. 351-352)

Thus Stevens and Volkman are led to hypothesise that jnd's based upon a physiological process involving a change in the general location of excitation will be found to be subjectively equal in size, whereas all jnd's produced by adding more excitation to excitation already present will be found to increase in subjective size with increased magnitude of the stimulus.

Although the mel scale is widely accepted as a standard pitch scale Gulick (1971) points out that halving the pitch is not equivalent to a step across a constant number of jnd's (p. 138) and he also points out that the mel scale is not identical with the cumulative jnd curve. Gulick has this to say about the mel scale in summary, p. 138:

"While the mel scale is of some value in showing a general relationship between pitch and frequency, some question as to its numerical properties can be raised. In our laboratory we have found only limited success with it because judgments of pitch are influenced to a great degree by the particular order of presentation of the standards selected. Furthermore, if mels are to represent a useful scale, then not only must $N/2$ mels be half the pitch of N mels, but N mels must be twice the pitch

of $N/2$ mels. This requirement is not always met by psychophysical data when different references are employed."

Thus although the scale proposed by Stevens and Volkman is widely used, it has its limitations. A very wide range of frequencies is dealt with on this scale, i.e. the total pitch range which is perceptible to the human ear. In discussions about the intonation of the human voice we are concerned with a very small proportion of these frequencies near the lower end of the scale. Since the mel scale has its shortcomings, and since it is generally agreed (by Gulick, Stevens and Volkman, etc.) that the lower the pitch the more closely it is correlatable with the frequency measurement, for the moment we seem to be justified in relating pitch to fundamental frequency alone rather than attempting to relate pitch to the mel scale.

2.4 Psychophysical and linguistic judgments

Most of the above investigations involving the physical and psychophysical correlates of stress and intonation have involved perceptual decisions of some kind (e.g. is this sound twice as loud as the other sound, etc.), but we have not yet discussed experiments involving linguistic judgments. This section, then, will examine some of the experiments dealing with inton-

ation which have been carried out involving linguistic judgments. There are many experiments in this field but we will divide them broadly into two main areas -- dealing with experiments which involve grammatical or semantico-syntactic judgments and dealing with experiments which involve attitudinal judgments.

2.4.1 Linguistic judgments of a semantico-syntactic nature

Hadding-Koch and Studdert Kennedy (1964, 1965) carried out a series of experiments using Swedish and American subjects to determine the relationship between psychophysical judgments and syntactico-semantic judgments. They used both two and three category tests asking subjects to judge whether an utterance was -- a) rising, falling or level and b) whether the utterance was a statement, question, or 'reflection'. I would like to refer to the latter b)-type experiments as experiments of a semantico-syntactic nature. Experiments involving judgments about whether utterances are questions or statements involve decisions as to the illocutionary function of the utterance, not necessarily referring to the syntactic form of the utterance.

Hadding-Koch and Studdert-Kennedy found that their third categories (level, and 'reflection' respectively) tended to 'nibble away' at the results of the two 'main' categories. They therefore set up a later experiment in 1971 (cf. Studdert-Kennedy and Hadding-Koch, 1973)

which only involved the two main categories at each level, i.e. terminally rising versus terminally falling contours, requiring psychophysical judgments, and contours functioning as questions or statements, requiring linguistic judgments. I will discuss this later experiment since it attempts to extend the earlier work.

This experiment as reported by Studdert-Kennedy and Hadding-Koch in 1973 deals exclusively with one of the acoustic correlates of intonation (i.e. fundamental frequency) since variations in fundamental frequency over time is said in the literature to be the strongest single cue to intonation (Bolinger, 1958; Denes, 1959; Fry, 1958; Lehiste, 1970). The experiment was initiated in order to examine the importance of the terminal glide in utterances, since the terminal glide is commonly said to be the acoustic cue for judging an utterance as a question or as a statement. The earlier experiments had shown that listeners frequently judged a falling glide as rising and a rising glide as falling but whether such judgments were due to psychophysical judgments or linguistic judgments was unknown. Studdert-Kennedy and Hadding Koch (1973) systematically manipulated the contour of an utterance by varying the fundamental frequency at the stress peak, at the 'turning point' before the terminal glide, and at the end point. Listeners were then asked to classify each contour as (1) question or statement (requiring a linguistic judgment) and (2) as

having a terminal rise or fall (requiring a psychophysical judgment). This experiment also included non-speech judgments of rising versus falling contours in order to find out the auditory capacity of listeners for judging the terminal glides of matched non-speech contours.

The utterance 'November' was spoken by an American male voice. F_0 was then manipulated over a range of 85 cps to 220 cps and the F_0 values at the most important points of the contours (i.e. starting point, peak, turning point and end point) were selected to represent four different F_0 levels. The four levels were based on a previous analysis of a long sample of speech by a speaker with this particular range (Hadding Koch, 1961, p. 110 ff.)

"The contours range between two poles that may be marked 2 44 3 and 2 11 1 . All contours start on a F_0 of 130 cps (level 2), sustained for 170 msec. over the first syllable. They then move, during 106 msec., to one of three peaks: 130 cps (L, or low, level 2), 160 cps (H, or high, level 3), 200 cps (S, or superhigh, level 4). They proceed during 127 msec., to one of four turning points: 100 cps (high level 1), 120 cps (level 2), 145 cps (low level 3), 180 cps (high level 3). Finally, they proceed, during 201 msec., to one of six-end-points: 85 cps (level 1), 100 cps (high level),

120 cps, 145 cps, and 220 cps (level 4). Peak, turning point, and end joint are each sustained for 32 msec. The combination of three peaks, four turning points, and six end points yielded 72 contours, each specified by a letter and two numbers (e.g., S24, L36) and each lasting 700 msec."

These contours were then presented to Swedish and American judges. The results did not differ greatly from one language group to the other and the main findings for the psychophysical judgments matched the findings for the linguistic judgments. The superhigh peak, even when followed by a high or moderately high turning point is accepted as a fall or statement provided the terminal fall is large enough. The lower turning point (i.e. the larger the fall from the peak), the less the required terminal fall. On the other hand, some terminally level contours and even terminally rising contours are also accepted as statements. Evidently the terminal fall is not essential if preceding sections of the contour are low enough, or are falling from a moderate level. Broadly speaking then peak, turning point and terminal glides interact to produce the contour of an acceptable statement-type, which has a low to high (rarely, and only for U.S. superhigh) peak and is latterly low, falling, or both.

The range of preferred question contours includes

the expected continuously rising contour of American English and the superhigh peak contour of Swedish, but other contours were also accepted. Initially low and falling contours were heard as questions if the terminal rise was large enough, and even a terminally level contour received more than 80% question judgments from both groups when the preceding section of the contour had been steadily rising. Thus a generally accepted question displayed either a rise from peak to turning point ~~or~~ a relatively large terminal rise.

Katherine Mickey (1977) points out that such a procedure as the one used by Studdert-Kennedy and Hadding Koch is a very common one in this field (see also experiments by Majewski and Blasdel, 1969; and Uldall, 1962). In all of the above experiments subjects have been asked to classify utterances which varied in fundamental frequency contour as 'question' or 'statement' (in Uldall's experiments, 'question' and 'statement' were the extremes of a seven-point scale). The results are commonly presented in graph form with the percentage of subjects responding 'question' constituting the Y variable, and the numerical value of the varying acoustic parameter as the independent X variable. Mickey points out that:

"In such cases, choosing the X variable may present problems. For example, in constructing a set of

stimuli which vary in extent of terminal glide, one would probably also vary the end-point fundamental frequency -- and either parameter might be chosen as X. Studdert-Kennedy and Hadding choose extent of terminal glide (in c.p.s.), and conclude that the percentage of subjects responding 'question' depends on that parameter. It seems clear that it would be useful to find a means of assessing the relative merits of different choices of x."

(Mickey, 1977, p. 21)

Mickey also points out that more problems arise if the Y variable appears to depend on two or more parameters at once (e.g. peak Fo and extent of terminal glide). She then goes on to set up two experiments in order to determine whether speakers can distinguish different pitch levels for different speakers, since a man speaking 'high' may well be lower on an absolute scale than a child speaking 'low'. In both experiments, subjects were asked to distinguish 'high' and 'low' rising intonations in synthetic 'yes's' and 'no's'. Mickey decided to use variations of rising intonation since Gimson (1970) and O'Connor and Arnorld (1973) state that rising glides may either extend from low to mid signalling that the utterance is unfinished or continuative, or from mid to high signalling that the utterance is a form of question. They also note that other varia-

tions of starting and end points between low and high are also possible. By restricting the experiments to variations of rise Mickey dispenses with one variable on the acoustic parameter by keeping the direction of movement constant.

Within each set of stimuli, F_0 varied over a range lower than would be used by any particular real-life speaker. In experiment 1, there was a variation in formant frequency and F_0 level, but not in extent of rise. In experiment 2 the formant frequencies were fixed, but there was variation in both F_0 level and extent of rise. After the experiments had been performed the results were analysed using a statistical technique which made it possible to assess the relative merits of different choices of X variables, and, for a particular choice of X, to evaluate the relative importance of the single variables involved. The technique used was multiple linear regression analysis. This same technique was then used to analyse the published data from the question / statement experiments of Studdert-Kennedy and Hadding, and Majewski and Basdall.

When all the results from the above experiments were considered, the analysis indicated that the best choices of X were (1) log endpoint F_0 , and (2) sets of parameters which sum to this value. It is interesting to note at this point that in this synthetic speech situation subjects seemed to judge the stimuli as if

they came from the same speaker despite the fact that the range of Fo values spanned by the stimuli was larger than the range which would be used by a 'real' speaker. The results showed that the listeners evaluated the stimuli with the highest ending Fo as 'high' or 'question', and the stimuli with the lowest ending Fo as 'low' or 'statement'. The value of endpoint Fo for which 50% respond 'high' or 'question' was at the middle (on a logarithmic scale) of the total range of Fo values spanned by the stimuli. Mickey summarises the results as follows, p. 38:

"Considering the results of all the experiments together, we have come to favour the hypothesis that the Fo level at the end of an utterance is primarily responsible for conveying question /statement information (cf. Majewski and Blasdell, 1969; Uldall, 1962), while the contour shape conveys other information. In saying this, however, it should be remembered that, given the intonation patterns which occur in English, utterances with high endpoint Fo will generally end with large rising glides -- in the everyday speech situation, the two probably cannot be viewed as independent parameters."

Thus despite the fact that Mickey takes great care to eliminate as many variables as possible and uses a

statistical technique which can separate one parameter from another, her conclusions do not separate Fo terminal level from final terminal glide. Also, her results do not account for the conclusions of Studdert-Kennedy and Hadding given above, where the terminal endpoint can be over-ruled by the terminal glide.

All of the above experiments used single words as the basis for the psychophysical and 'linguistic' judgments (Studdert-Kennedy and Hadding used the word 'November'; Majewski and Blasdell used the word 'farmer'; and Mickey used the words 'yes' and 'no'.) Therefore I do not think any conclusion can really be drawn about the effect of the overall contour shape except when the overall contour is found on one word utterances.

The above experiments indicate only some of the problems involved when setting up psychophysical and linguistic experiments involving intonation. Many acoustic parameters are involved which are interrelated in such a way that it is very difficult to separate one from the other, also a greater variety of linguistic stimuli must be examined in order to reach conclusions about the linguistic functions of intonation, i.e. when examining the question / statement dichotomy, not only should we vary the acoustic parameters over one word, but also over long utterances. How do we know that one-word experiments produce valid 'linguistic'

results? The above experiments could also be carried out using single word stimuli in context as well as out of context. Kenworthy (1977) carried out such experiments with quite different results for one word stimuli out of context and one word stimuli in context. She found that when the stimulus was presented out of context listeners would base their judgments on acoustic bases such as have been mentioned previously (i.e. a high rise was judged to be a question), but when the same stimulus was presented in context, the context determined the results in such a way that a high rise was judged to be a realisation of listing in one particular instance.

Thus the status of the conclusions drawn by the above investigators must be made clear. Judgments may vary depending on the information available. When a stimulus is presented out of context we seem to make psychophysical judgments, when a stimulus is presented in context we make judgments which take the context into account. Cf. Wingfield (1975) where he says:

"phonemic judgments completed at one level, are modified, or even rejected, as analysis continues to clausal and sentential levels. Demand characteristics of experiments which terminate processing at the level of phonemes, syllable or words, for example, thus yield rules valid only for these

levels. Later levels of processing include review of all evidence, and may perhaps reverse earlier decisions in the light of further information."

(p. 152)

2.4.2 Judgments of an attitudinal nature

"Attitudinal nature" in this section is being used to refer to experiments involving judgments about the attitudinal content of an utterance or the emotional content of an utterance. There are many of these experiments in the literature. We will examine only some of them.

The experiments carried out in the area specified in this section i.e. involving attitudinal or emotional judgments) may take several different forms. The experiments may involve open-ended judgments where subjects are asked to label a stimulus with no restrictions placed on the number of labels available; the experiments may take the form of forced choice judgments, where the subject is presented with a specific number of labels, and has to match the stimulus against one of the restricted number of options open to him; or the experiment may take the form of same / different judgments, where the subject is presented with a model stimulus and is then asked whether the variable stimulus is the same as the model or different. These three methods are the most common of the methods used in at-

titudinal experiments.

These various methods have their intrinsic shortcomings. The open-ended experiments can often result in there being as many labels as there are stimuli; the forced-choice experiments involve the investigator's subjective choice of labels which need not necessarily be the most appropriate for the experiment; in the model experiments, the investigator cannot be sure whether the subjects are making judgments according to the label of the stimulus model, or according to the acoustic or perceptual parameters involved.

One experiment set up by Lieberman and Michaels (1962) examines the contribution of the two acoustic parameters (fundamental frequency and amplitude) to the transmission of the emotional content of speech. Synthesised speech was used for this experiment based on recordings made for a previous experiment (Lieberman, 1961). This previous experiment required speakers to read sentences out of context as though they belonged to one of the following eight emotional categories: (1) a bored statement, (2) a confidential communication, (3) a question expressing disbelief or doubt, (4) a message expressing fear, (5) a message expressing happiness, (6) an objective question, (7) an objective statement, and (8) a pompous statement. The readings of these sentences were then categorised by a group of linguistically naive listeners in a forced judgment experiment to

select the most identifiable utterance from each of the eight emotional categories for each sentence. A panel of trained observers then listened to the same set and rejected those utterances which they found unnatural or strained. Such procedures are often used when setting up the categories for a forced judgment test or rather the exponents of those categories which a 'trained group of people' decide are the best realisations of these categories. Under such circumstances what does 'trained' mean? If it means that the 'trained' observers have been taught what the 'best' realisation of certain emotional categories are, then the decisions are primarily due to the 'trainer' (who often happens to be the investigator in charge of the experiment). There does not seem to be any clear-cut way of increasing the objectivity of such experiments. Whenever naive subjects are asked to judge utterances 'objectively', the results are frequently widely divergent even amongst a small number of judges (cf. Fairbanks and Provonost, 1939).

Nevertheless, utterances for the Lieberman and Michaels experiment were categorised in terms of the eight emotional categories mentioned above, then the frequency and amplitude information contained in the utterances was isolated.

The results of the experiments showed that there was no one single acoustic correlate of the emotional

categories. Lieberman and Michael summarise the results as follows:

"Phonetic content, gross changes in fundamental frequency, the fine structure of the fundamental frequency, and the speech envelope amplitude, in that order, all contributed to the transmission of the emotional modes." (1962, p. 927)

They also note that the different emotional categories did not all depend on all the acoustic parameters to the same degree, and also that different speakers or listeners preferred different acoustic parameters for the transmission of the same emotional category. The results of the experiments showed that with unprocessed speech the listeners were able to identify the emotional content correctly 85% of the time; when only pitch information was presented, correct identification was made 44% of the time; when amplitude information was added to the pitch information, identification rose to 47%. A 120 cps monotone with amplitude information resulted in only 14% identification. From these results Lieberman and Michaels conclude that the fine structure of the fundamental frequency, i.e. the perturbations in fundamental frequency appears to be an acoustic correlate of the emotional modes. But, as has been pointed out in Chapter 1, there are many other features

of the acoustic wave form which have not been taken into account such that any one of these features or any combination, or the total combination, may be the acoustic correlate of the emotional modes.

Let us investigate the various 'emotional' categories which have been used in such experiments. One of the widest range of categories used in a forced choice experiment is found in Uldall (1964). Uldall presents 28 opposed adjectives with a seven place scale between each opposed pair as follows:

bored	— — — — —	interested
polite	— — — — —	rude
timid	— — — — —	confident
sincere	— — — — —	insincere
tense	— — — — —	relaxed
disapproving	— — — — —	approving
deferential	— — — — —	arrogant
impatient	— — — — —	patient
emphatic	— — — — —	unemphatic
agreeable	— — — — —	disagreeable
authoritative	— — — — —	submissive
unpleasant	— — — — —	pleasant
genuine	— — — — —	pretended
weak	— — — — —	strong

In Uldall's experiment, therefore, there are actually (14 x 7) or 98 possible categories, plus all the possible combinations of categories which were available for combination, i.e. although all 98 categories were available, each subject would only select one place between each opposed pair (I assume, although this is not made explicit), therefore not all possible combinations were available).

Four sentence types (of the syntactic form statement, yes / no question, question-word question, and command) were then uttered on a rising intonation and 16 intonation contours were imposed on these utterances varying the fundamental frequency synthetically as follows: pitch range was either wide or narrow, terminal end-point was either high, mid or low, and the shape of the contour was either in the form of one direction of movement or with a change of movement.

The contours which were judged to be the most neutral for each of the sentence types were as follows:

Statement:	final rise ending mid
Yes/No question:	final rise ending high
Question word question:	final rise ending high or mid
Command:	final rise ending high or mid

Since all of the original utterances had been uttered on a rise, it is perhaps not surprising to find

that no neutral judgments were given for any contour ending on low (formant frequencies for example might make any contour not uttered on a rise sound non-neutral). The question-word question and the command sentence types appear to have variable intonation realisation, but the question / statement findings seem to support Mickey's findings (Mickey also used original utterances spoken on a rise and varied the end-points). Other general conclusions reached by Uldall in this experiment are stated as follows:

"The more 'lively' the contour is, the more stable is its position in the 'semantic space' over the different sentences. Contours involving wide range and a change of direction, always occupy the same sector of the space, the pleasant, authoritative, strong one...

The less 'interesting' the intonation contour is, the more influential the sentence itself is in the judgment of the total effect, and vice versa."
(in 1972 reprint, p. 258)

Thus despite the great variety of category choices available, or perhaps because of it, no strong conclusion can be drawn from this experiment except to say that the wider a contour is, the more 'positive' it seems to be.

We can see from the above experiments that forced-

choice category experiments may range in number of categories from 2 to 98, yet still produce similar results given similar stimuli. The categories vary widely from one experiment to another, yet even when the same categories are used the results may completely contradict each other. For example Garding and Abramson (1965) find that a neutral statement is realised by a fall from mid, whereas Uldall (1962) finds that a neutral statement is realised by a final rise to mid. The reasons for such conflicting results are many and varied: the stimuli may vary from one experiment to another; the categories vary from one experiment to another; and the acoustic parameters being investigated may vary from one experiment to another.

Let us examine two experiments to see how these parameters may vary from one to the other. We will compare Faibanks and Provonost (F & P), 1939 with Garding and Abramson (G & A), 1965.

(G & A) set up 5 categories by asking informants to read a sentence expressing:

1. Neutral statement
2. Yes or no question
3. Anger
4. Delighted surprise
5. Counting in a series

These sentences were read out of context. (F & P) also set up 5 categories, these were as follows:

1. Contempt
2. Anger
3. Fear
4. Grief
5. Indifference

A test passage was inserted into different context passages each of which 'favoured intense expression of one of the above emotions' (p. 88). Thus in (G & A)'s experiment a sentence was read out of context, whereas in (F & P)'s experiment a passage of about 3 sentences was read in context. The two experiments differ already in the method of eliciting the stimuli required for the experiment. The categories used in each experiment are different apart from one (anger), thus another variable is involved. Both experiments then required subjects to judge the appropriateness of the read sentences as realisations of the categories, then in both experiments, the investigators took the final decisions as to which sentences were most appropriate realisations of each category.

(F & P) then continue by examining the pitch characteristics of the contour which receives the maximum 'correct' identification for each category, i.e. forced choice testing is continued and the contour which is

identified with a particular category most often is then examined in detail with specific reference to the F_0 measurements.

(G & A), on the other hand, produce synthetic contours using the 'best' real speech contours as the carrier sentences. The fundamental frequency characteristics were then varied on the synthetic sentences, and subjects were asked to make same / different judgments about the sentences in the following way: a model contour (contour A) is presented to the subject, then a series of stimuli are presented, and the subject ranks the stimuli according to whether the stimulus is like or unlike the model. Thus (G & A) were able to test the autonomy of the contour unit which represented each category by testing how much variety was permitted within a category, and by testing when one contour was judged to belong to more than one category (G & A) could discover where the limitations of each contour were specified.

Thus each of the two experiments used a different method of testing (G & A using a same / different technique, whereas F & P use a category forced technique), and they are also testing different parameters in that (G & A) is testing how much a contour can vary within a given category whereas (F & P) are essentially concerned with measuring the F_0 characteristic of the 'best' realisation of a category.

Let us compare the final findings for the one category that these experiments have in common, i.e. Anger.

Garding and Abramson summarise their results as follows:

"Shape: A falling pitch movement from a high level.

Size of interval: From 70 to at least 120 cps. The end point can be raised 30 cps (Contours 13, 14, 15, Test 6, p. 76). The beginning point can be moved up at least 20 cps (Contours 9, 10, Test 6).

Level: The level from which C starts is more important than the interval (cf. Contour A).

Changes in shape: A level of rising beginning which occupies one fourth of the duration of the final syllable above C can be omitted (Contours 18 and 19, Test 6). Even if this initial rise starts from a much lower level (e.g. at 130 cps as in Contour 16, Test 6, p. 76) the contour is still identified with C.¹ It should be noted that a similar change in A takes it out of the A-category. We can conclude that C has a larger range than A. The range of a contour is probably proportional to the amount of pitch movement that is involved in it." (1965: 72-3)

Fairbanks and Provonost summarise their results as follows:

"Anger. (1) The widest mean phonational range. (2)

The widest mean extent of inflections in all types and extents of inflections, this increase being most marked in upward inflections. (3) The most rapid rate of pitch change during inflections. (4) The widest mean extent of all pitch shifts (together with Fear) and of downward pitch shifts within phrases. (5) The greatest number per second, during phonation only and during the total speaking time, of pitch changes one semi-tone or greater in extent. (6) The smallest number per second during phonation of changes less than one semi-tone in extent. (7) A greater number per second during phonation of changes one semi-tone or greater in extent than of smaller changes, the reverse being true for other emotions." (1939: p. 103)

From the above discussion then, it should be clear that experiments involving judgments of emotional categories and their realisation in acoustic terms, can vary so widely that it is very difficult to even begin to compare the results of one experiment with the results of another.

For these reasons we shall ignore the 'emotional' or 'attitudinal' function of intonation and concentrate on the 'linguistic' function of intonation.

Chapter 3.

In this chapter I would like to discuss in detail two of the main systems of intonation which have been proposed for British English. The first of these systems is the one defined by Halliday (1963 and 1970), and the second is Crystal's system as defined in 1969. I will deal briefly with these authors' respective views on the relationship of the intonation system to other linguistic systems, but I will concentrate my main discussion on some of the details of the working out of the intonation system itself.

3.1 Halliday's system of intonation

Let me start with Halliday's theory as presented in 1963 and 1970. I am concentrating the discussion on these two works since they constitute the most detailed of his works on the system of intonation itself and its realisations.

Throughout Halliday's works on intonation he makes it quite clear that intonation functions as part of the grammar system, cf. Halliday (1963), p. 3:

"English intonation contrasts are grammatical:
they are exploited in the grammar of the language".

But Halliday's conception of 'grammar' is perhaps a little different from a standard or conservative con-

ception of grammar. By 'grammar' Halliday does not mean purely syntax. In 1970 Halliday states the relationship between grammar and intonation very clearly:

"..there are always various possible intonation patterns; and all these will carry different meanings. These different meanings are part of English grammar... the distinctions expressed by the choice of different tones are also distinctions in meaning, and they are the same general kind (as grammatical distinctions such as different types of subordinate clauses); so they too belong in the grammar (and, within grammar, the realm of syntax). Intonation is one of the many kinds of recourses that are available in the language for making meaningful distinctions". (1970, p. 21)

This view might be considered to be fairly unusual. Most theories of intonation believe either that intonation functions syntactically (cf. Trager & Smith, 1951) or that intonation functions semantically by signalling attitudinal contrasts (cf. O'Connor and Arnold, 1973; Fairbanks and Provonost, 1938; Uldall, 1964, etc). Halliday believes that intonation sets up meaningful contrasts which operate in the grammar of a language.

The problem then is, how does Halliday

relate his various systems to each other? How does the system of syntax interact with the system of semantics and the system of phonology? How are these 'systems' related to 'levels' of grammar? Does Halliday believe that any set of meaningful contrasts constitutes a system, and that each system then constitutes a 'level of the grammar'?

In 1963 Halliday says:

"Whenever we describe a language we are concerned with meaning, and all contrast in meaning can be stated either in grammar or in lexis...there are only two kinds of formal pattern: grammatical and lexical" (p. 3)

Since intonation contrasts are not lexical, they must be grammatical. Halliday goes on to say:

"therefore in a description of the grammar of spoken English, "intonational" and "non-intonational" systems figure side by side. They are not to be treated as systems of different types. Moreover since "intonational" systems operate at many different places in the grammar, they will not be isolated in a chapter by themselves, but incorporated throughout the description wherever appropriate." (p. 4)

We seem to have a situation where there are several independent systems which have been set up according to contrastive sets of some kind, each system interacting with the other. Both Halliday and Crystal talk about systems in such a way that one system can 'contain' another system or set of systems. For example, we can have a 'macrosystem' such as the prosodic systems proposed by Crystal which contain other 'systems' such as the voice quality system and the intonation system. The intonation system can then include 'micro-systems' such as the tone system and the pretonic system, and so on. Throughout this thesis I will not specify whether systems are 'macro' or 'micro', rather the attitudes of Crystal and Halliday will be adopted, i.e. a system is a contrastive set of items, and each item in a system may in its turn be realised by a contrastive set of items, thus systems may be contained within systems.

Any independent set of contrastive features, therefore, will be said to constitute a system.

There are no direct correlations between one system and another for Halliday, i.e. he does not want to set up a system which would assign direct phonological exponents to all 'grammatical' categories since, he says, there is no language in which all classes at every rank are marked phonologically.

Despite the fact that Halliday says that intonational systems "operate at many different places in the gram-

mar" he only ever relates the intonational system to his system of 'mood' of 1963, p. 21 where he "ranges the tones in grammatically contrastive sets" starting with the "clause system of mood, since this determines which tone is neutral in a given instance". It is unclear just exactly what Halliday means by this last sentence. He cannot mean that each term in the mood system is uniquely correlatable with a different term in the tone system such that tone 5 is neutral for 'affirmative', for example, since he says:

"tone 1 can be regarded as neutral for all terms in the mood system except polar interrogatives".
(1963, p. 22)

What Halliday seems to be doing instead is setting up a system of contrasts called the 'mood' system, setting up a system of contrasts called the 'tone' system, and then examining the result of various combinations of the two independent systems. Let us set up a table showing the interaction of these two systems:

TONES	AFFIRMATIVE		INTERROGATIVE		IMPERATIVE	MOODLESS
			POLAR	NON-POLAR		
1	statement or answer NEUTRAL		DEMAND question	question NEUTRAL	command, answer, exclam- NEUTRAL ation	
2	statement or answer CONTRADICTION		question NEUTRAL	question or question (or with negative ECHO QUESTION CONTRADICTION answer or NEGATIVE command)		
3	statement DEPENDENT OR CONFIRMATORY		answer NON-COMMITTAL		answer NON-COMMIT- TAL	
4	statement or answer RESERVATION				statement or answer RESERVATION	
5	statement UNCONDITIONAL INDEPENDENT				answer COMMITTAL	
13	statement NEUTRAL					
53	statement COMMITTAL					

As can be seen from the above table the full range of co-occurrences were not explored by Halliday.⁽¹⁾

Halliday defines the system of mood as follows:

affirmative clause	(subject before predicator)
interrogative clause	(subject after first word of predicator)
imperative clause	(with predicator but no subject)
moodless clause	(with no predicator)

Halliday does not define predicator here, but presumably this term refers to verb or verb phrase. The above terms, then refer to the four possibilities in Halliday's clause system of 'mood', and constitute a primary system at the first degree of delicacy (p. 21). It should be noted that the definitions of these terms in the mood system are very simplistic and all consist of linear constraints referring purely to the ordering of the lexical items at the surface structure. Since Halliday really only discusses intonation in relation to the mood system, and does not mention other systems which might co-occur with intonation, it is not clear whether he would wish to relate the intonation system to 'deep structure syntax' (for the purposes of disambiguation for example) as would Sledd (1962), Stockwell (1962), Lieberman (1967). In the works mentioned

so far, Halliday does not maintain a deep / surface opposition, but prefers to treat all intonational and non-intonational systems side by side (p. 3-4)

If we examine Fig. 1 in detail we find that Polar Interrogatives seem to stand out from the other terms in the mood system as being different. All the other terms are neutral when co-occurring with tone 1, non-committal with tone 3, reserved with tone 4, and committed with tone 5. Since Halliday would like to maintain that intonation is no different from any other system, it is surely interesting to note that any term in the tone system can co-occur with any term in the mood system, but that whenever a specific tone appears there is a striking consistency in 'meaning' associated with it. For example tone 1 is used consistently as the neutral tone. The system of tone, then, is obviously not set up according to contrastive functions operating in the mood system.

If the same tone occurs neutrally for all Halliday's systems of mood, how then does he set up a system of tones which function contrastively, since a particular tone is obviously not a realisation of a particular term in the mood system. Halliday refuses to set up a system based upon observations of "like" versus "unlike" alone, since he says the result is likely to be either excessively complicated or distorted. Nevertheless he does set up such units initially (based on "like" versus

"unlike") but he then asks himself whether a contrast between two such units is meaningful:

"Is this distinction, which I can abstract from observations of the substance, meaningful?"

(1963, p. 5)

Halliday's system of tone contrast, then, is determined according to the 'meaning' of the tone unit. This methodology is dubious for several reasons, some of which were mentioned in chapter 2. For example one must distinguish between categorisation and discrimination, i.e. does one wish to know if tone 1 can be discriminated from tone 2, or does one wish to identify 5 categories of tone and fit the realisations into the categories. Under optimal conditions the threshold for detecting a difference in frequency between two successively presented sinusoidal tones is of the order of 1 part in a 1000 (Rosenblith and Stevens, 1952). On the basis of comparative judgments such as these, human listeners can distinguish about 350,000 different pure tones (Stevens and Davis, 1938). But note the relative inability of a hearer to identify a name (i.e. categorise sounds) presented for individual identification. Pollack (1952) showed that listeners could not consistently identify more than 4 or 5 pure tones when each tone was presented individually for identification. There is

therefore a gross disparity between these performances of two different tasks.

The method of contrasting tone 1 with tone 2 when they are produced on the same lexical content can be compared with contrasting cardinal vowel 1 with cardinal vowel 2, the principal test being to see whether the Hearer can perceive a difference between the two. If the Hearer can perceive a difference, the next question to be answered is, is the difference functional / meaningful? At least when the Hearer is comparing two cardinal vowels, if one is substituted for the other the difference is meaningful when another lexical item is perceived. But when two tones are substituted for each other they are said to be functionally contrastive if there is a difference in meaning. How is this difference to be determined? If we ask a Hearer to describe the difference in meaning, how do we delimit the nuances of meaning which can be read into very slight differences of intonation? The possible meaning categories associated with tonal contrasts seem endless (cf. section 2.4.2). On the other hand if boxes are provided for Hearers to put the tunes into, who determines which boxes are to be made available? The labelling of the boxes will to a very large extent determine the results. (cf. the various experiments described in 2.4.2).

The above criticisms refer only to ad hoc examples where the same lexical content is produced with differ-

ent intonation contours, many more problems arise if the lexical content is allowed to vary. In the latter case the context (and the lexical items themselves) will to a large extent determine the 'meaning' of the tone (cf. Brazil 1978). For example the lexical content in the following examples seems to me to be sufficient to account for the difference in meaning without even examining the context. Halliday, however, attributes these differences in meaning to the difference in tone, cf.

Tone 3: non-committal answer - //3 six foot//

3 I don't/know//

Tone 5: commital answer - //5 I certainly/do//

- //5 oh I'm / sure it /
was//

(Halliday, 1963, p. 24)

There are many problems involved in associating a particular tone with a particular meaning. Lexical content, context and the subjective labelling of the perceived differences must all be taken into account when a particular meaning is associated with a particular utterance.

Nevertheless Halliday does set up a system of meaningful contrasts which are realised by different tones. Halliday says there are three distinct meaningful choices

involved in setting up such a system and these are as follows:

"first, the distribution into tone groups -- the number and location of the tone group boundaries; second, the placing of the tonic syllable -- the location, in each tone group, of the pretonic and tonic sections; third, the choice of primary and secondary tone .. The three selections are independent of one another." (1963, p. 15)

We are therefore presented with three systems of contrast operating (apparently) independently. But according to Halliday's definition of the terms, the systems cannot be entirely independent. The first two systems in fact are so closely related they are inseparable. Halliday says:

"any one element may be assigned a tonic and therefore demand a new tone group" (1963, p. 17)

Thus it would seem to be the case that the selection of the tonic determines the domain of the tone group, thus implying that the boundaries of the tone group depend to some extent on the identification of the tonic. Yet, on the other hand the definition of the tone group and its boundaries is said to determine the

tonic, cf. p. 18:

"the choice of how many tone groups, and where their boundaries are, goes a long way towards determining the choice of how many tonics, and where they are located."

A certain amount of circularity seems to be involved here. We are given criteria for identifying tonics (a tonic being defined as the most prominent syllable in the tone group) but we are told on p. 14 in a footnote that "the location of the tone group boundary is a theoretical decision". How then, are we to determine the tonic if we cannot decide where the tone group boundaries are located, and if we are unable to identify tonics, we cannot identify the domain of the tone group.

This is a fundamental problem which undermines the principles involved in setting up Halliday's system of contrastive intonation and this problem is not resolved in his later works.

Given that we can identify tone groups in some unspecified way, Halliday's contrastive systems operate within the tone group as follows:

"The tone group comprises two elements of structure... "tonic" and "pretonic". The element "ton-

ic" is obligatory: it is present in every tone group. If the pretonic is present it always precedes the tonic... All primary contrasts are carried by the tonic, but some secondary contrasts are carried independently by an element preceding the tonic." (1963, p. 6-7)

It seems incongruous to allow a system of secondary contrasts on elements preceding the tonic while disallowing any contrasts following the tonic. The tonic is allowed to occur anywhere within the tone group, cf.

"the tonic may start at the beginning of ... any strong syllable." (1963, p. 10)

Therefore it would surely be reasonable^a to expect secondary contrast also to occur anywhere within the tone group, either preceding or following the element of primary contrast, yet Halliday says:

"the pitch movement is distributed over the whole tonic, and there is no post-tonic element ... in no case is a separate contrast carried by any feet after the tonic foot." (1963, p. 8)

Halliday does allow feet containing 'strong' or 'stressed' syllables to occur both before and after the

tonic foot but no system of secondary contrast is carried by feet following the tonic while a system of secondary contrasts is said to operate on feet preceding the tonic foot. This seems to me to be an inconsistent way of dealing with strong or stressed syllables which obscures possible generalisations. If a system of secondary contrasts was allowed to operate after the tonic element as well as before the tonic element, Halliday's system of double tonics could be accounted for in a way which would conform with the rest of Halliday's theory rather than seeming to contradict it. As the system stands, when there is a double tonic the first element is said to carry the primary contrast while the second element carries the secondary contrast. This position runs counter to all Halliday's previous statements about there being no secondary system operating after the tonic element. If secondary contrasts were allowed to follow the tonic element regularly, then double tonics would not weaken Halliday's theory but rather support a more symmetrical system which would account for more of the data without seeming to consist of contradictions.

Let us now examine the system of secondary contrasts which operates both at the tonic and at the pretonic. These systems are said to be independent from each other, but they are not entirely, cf. the following quotations from Halliday:

"In general, the neutral pretonic remains fairly level before the neutral tonic, works up towards the 1 + tonic and down towards the 1-"

(1963, p. 11)

The pretonic in general, therefore, seems to be determined according to the tonic following it (or perhaps the pretonic determines the choice of tonic), i.e. Halliday seems to be saying either that when there is a 1 + tonic the neutral pretonic rises, or when there is a rising pretonic, there will be a 1 + tonic. When there seems to be a true choice of pretonic, e.g. when one can choose to select either the 'neutral' or the 'listing' pretonic, the 'sense' seems to determine which pretonic should be used, i.e. the 'listing' pretonic occurs only when "one is enumerating a list of things or of people" (1970, p. 14).

Since Halliday believes that the pretonic is in general determined by the tonic it might be more sensible to look at the tone group as a whole in terms of the total pitch movement in order to truly establish to what extent the two systems are bound together. If the unit was examined as a whole the tone group might appear to be a much more autonomous unit, whereas at the present time it seems to be something which is very difficult to delimit and which is composed of bits and pieces of different systems which sometimes operate in

a contrative way but not always. If the tone unit were examined as a whole then, the tone unit as a whole could perhaps be seen to operate as part of a single, contrastive system.

Halliday's primary tone system is composed of 5 tones and two double tonics making up 7 primary tone contrasts. There are several anomalies within the system, however. We have already mentioned the oddity of the double tonics, but within the 5 primary tones there also seem to be inconsistencies. Tone 2 represents both a rising tone and a falling-rising tone, whereas tone 1 represents a falling tone, and a separate tone, tone 5 represents a rising-falling tone. If a generalisation has been drawn within tone 2 such that a rise is equated in some way with a fall-rise (both forms terminate at high and involve pitch movement from low to high) why should this generalisation not be extended such that a fall is equated with a rise-fall. Both forms terminate at low and involve pitch movement from high to low. If one pair of forms has been collapsed into one tone then surely an equivalent pair of forms should also be collapsed into a single tone. No reason is given for the different treatment of these apparently equivalent pairs of forms. The treatment of tone 2 begs yet another question. If the rise and fall-rise in tone 2 are treated as exponents of the same tone ending at high, why shouldn't tones 3 and 4 be subsumed under the same

heading since one is a rise and the other a fall-rise, both ending at mid?

Let us return to the problem of the double tonics. We have already proposed that since the initial tone is the primary contrast and the final tone is a secondary contrast (the secondary 'contrast' being entirely predictable as tone 3) the double tonic should not be a special instance within the system but should be a form of tone group where one allows for secondary contrasts after the tonic syllable. Since the secondary contrast seems to be entirely predictable the double tonics disappear as contrastive tones within the system since the contrast is carried by the initial element only (realised as one of the 5 primary tones, normally tone 1 or tone 5). An alternative proposal could be set up, however. If we extend Halliday's notion that the contrastive pitch movement is carried over everything^h following the tonic to include the tone group as a whole such that the contour of the tone group can be said to be contrastive, then tone 13 can be seen to be very similar indeed to tone 4, the difference being that tone 13 allows 2 prominent syllables within the tone group with the pitch movement divided between the two prominent syllables, whereas tone 4 allows only one prominent syllable within the tone group with all the pitch movement realised on that syllable, or extended to the end of the tone-group (cf. Collier, 1977, who found that

listeners grouped both tones 1 + 3 and 4 as members of the same group). Halliday already allows for a strong tonic and strong non-tonic, so there is no reason why this notion cannot be extended to allow for more than one prominent syllable within the tone group consistently. Halliday contradicts himself by insisting on only one prominent syllable per tone group in one breath then allowing double tonics and strong non-tonics in the next breath. Surely a greater generalisation would be captured if we consistently allow for more than one prominent syllable within the tone group and examine the pitch movement of the tone group as a whole.

Since Halliday takes a single grammatical category to demonstrate the different functions of contrasting tones within that category, the tones cannot be said to be differentiating between grammatical categories. Halliday summarises the functions as expressed by different tones as follows:

"If polarity is certain, the pitch of the tonic falls; if uncertain, it rises. Thus tone 1 is an assertion, or a query involving polarity, and tone 4, which falls and then rises, is an assertion which involves or entails some query. Tone 2 is a query, 2 being a query about a specific assertion; and tone 5, which rises and then falls, is a dismissed query, one countered by an assertion. Tone 3 avoids a decision;

as an assertion it is at best confirmatory, contingent or immaterial." (p. 28)

These statements concerning the function of contrastive tones do not indicate a syntactic function of intonation in the formal sense, i.e. tone 1 is not associated purely with declarative form, tone 2 with interrogative etc. in fact on the contrary tone 1 is said to be the neutral tone for all of these grammatical forms except polar interrogatives. The functions, as Halliday defines them, are related rather to the speakers' strategies involved in *the* discourse. The 'pure' system of tone seeming to consist of the contrast between 'certain = fall'; 'uncertain = rise'. In order to study such strategies fully, the context surrounding the utterance and the lexical content of the utterance itself must be taken into consideration together with the form of the intonation contour. Only then can we appreciate the contrastive function which can be identified within each system, before investigating the functions of intonation operating across various systems of contrast operating in the 'grammar'.

We must not only delimit the systems within which intonation is said to operate contrastively, we must also decide where intonation as a system belongs, what its sub-systems consist of, and how this system is related to other suprasegmental systems of speech. Let us now

turn to Crystal's theory of intonation.

3.2 Crystal's system of intonation

Crystal describes two of the systems of suprasegmental speech as the prosodic system and paralinguistic system, and he differentiates between these two systems on two levels -- the phonetic level and the functional level. Let us first examine the defining features of these systems at the phonetic level.

Crystal defines the system of prosodic features as being those "vocal effects constituted by variations along the parameters of pitch, loudness, duration and silence", saying that the above definition "excludes vocal effects which are primarily the result of physiological mechanisms other than the vocal cords, such as the direct workings of the pharyngeal, oral or nasal cavities: these are referred to as paralinguistic features." Crystal, therefore, differentiates between these two systems on purely physiological grounds. The differentiating features mentioned above, however, are very general and we must decide whether the definitions remain distinct when we examine the systems themselves. There is one general problem which arises immediately when we begin to examine the two systems individually. Crystal has defined the prosodic system as being constituted of variations along the parameters of pitch, loudness, duration and silence, but later on he states that

all paralinguistic features must, by necessity, possess the features of pitch, loudness, duration and silence to a certain extent since all speech does possess these features. These features must therefore, have variable status or as Crystal puts it (p. 139), "these (features) are variable in respect of their identification". This phrase is in fact the only phrase he uses to describe the relationship between the status of the above features as members of the prosodic system and their status as members of the paralinguistic system. If we look for further clarification of this relationship we find, in Crystal's concluding remarks almost exactly the same phrase being used:

"In summary, paralinguistic features may be described as combinations of physiologically grounded parameters with pitch, loudness, duration and silence being variable in relation to their identification."
(1969, p. 139)

If I understand the above somewhat obscure statement correctly, then Crystal seems to be saying that those features which make up the prosodic system are also part of the paralinguistic system and that the relationship between the status of these features in the 2 different systems is that they can be identified as belonging to either. If the above statement is correct

then the relationship between these two systems becomes very hazy indeed, such that one system seems to become almost part of the other so much so in fact, that there may be no necessity for keeping the two systems apart.

Let us look briefly at the features which realise these two systems. If we examine the parametric analysis of Crystal's paralinguistic features on p. 134 we find that at least 5 out of the 11 defining parameters refer exclusively to the state of the vocal cords. Thus the paralinguistic system not only includes variations along the parameters of duration, pitch, loudness and silence which are the defining features of the prosodic system, but it is also the case that almost half of the parameters which define the paralinguistic features refer exclusively to the vocal cords. Both of these facts seem to contradict Crystal's initial position and to obscure any phonetic basis for differentiating between these two systems, i.e. the defining features of these two systems overlap to such an extent that it is difficult to differentiate between them.

Crystal states (p. 138) that prosodic features normally co-occur with paralinguistic features, elsewhere he says that paralinguistic features are often defined in terms of prosodic features. The paralinguistic system is made up of features operating on a continuum which contains a neutral area (designated the norm)

which is said to characterise 'non-paralinguistic' speech. Because of the very close relationship between these two systems and the considerable overlap of defining phonetic features it might be reasonable to propose that the prosodic system is in fact the 'neutral' system which operates in that area of the continuum of the paralinguistic system which is designated as the norm. The two systems would then be collapsed into one with the prosodic system operating in non-paralinguistic speech and of necessity being present (in either extended or reduced form) in paralinguistic speech.

Let us now examine the functional distinctions between the two systems. Crystal states (on p. 130) "From the point of view of the primary semantic role of prosodic and paralinguistic features - the signalling of attitude - there seems to be little difference." On p. 129, however, the two systems are distinguished from each other as follows:

"At the 'most linguistic' extreme would be placed those prosodic features of utterance, describable in terms of closed systems of contrasts which are relatively easily integrated with other aspects of linguistic structure, particularly grammar, and which are very frequent in connected speech -- the variations in pitch and loudness ... under the general headings of intonation and stress respect-

ively. At the other 'least linguistic' end would be placed those paralinguistic features of utterance which seem to have little potential for entering into systemic relationships -- and are very infrequent in connected speech." (1969, p. 129)

Again we have the defining features set up on a continuum basis, which seems to me to draw these two suprasegmental systems closer together still. There are two features which seem to be the main distinguishing features between the two systems. Let us first examine the continuum with the two polars 'least linguistic' and 'most linguistic'. The two systems are distinguished along this continuum as being polar in terms of their linguistic function such that the prosodic system is at the 'most linguistic' extreme with the features "describable in terms of closed systems of contrasts which are relatively easily integrated with other aspects of linguistic structure ... whereas those paralinguistic features which have little potential for entering into systemic relationships and which have relatively little integrability for entering into systemic relationships" would be placed at the least linguistic end of the scale. The difference between the two systems then would seem to be scalar, based on the 'degree of linguisticness' of the system, but how are we to define one system as being 'more linguistic' than another sys-

tem? Is the 'more linguistic' system to have more communicative value? to have a more discrete set of contrasts? be more economic? to capture more generalisations? to attain greater observational, descriptive, and / or explanatory adequacy? Has the 'more linguistic' system to possess all of the preceding characteristics or only some of them? and if only some, then which ones? Crystal goes no further in defining how the degree of 'linguisticness' is to be evaluated therefore we have no means of differentiating between the two systems functionally.

One main feature of overlap between the two systems is exemplified by the independent system of Tension which belongs to both the prosodic system and the paralinguistic system according to Crystal. There are other instances of overlap too, such that one term in the paralinguistic system of voice qualifiers (creak) frequently co-occurs with one of the most linguistic terms in the prosodic system (pitch direction) (low fall). This is surely an unnaturally close relationship between two terms which are placed very far apart on the 'linguistic' scale. On p. 137, Crystal notes that the term low fall may be replaced by whisper or creak in the paralinguistic system without any difference in meaning such that these terms, from two apparently unrelated systems, may seem to have an identical function (this latter example is supported by Bolinger 1945, and Charleston, 1960).

The continuum of 'linguisticness', then, is not only very difficult to define but involves many inconsistencies and too much overlap.

The second feature which Crystal points out as a distinguishing feature is the 'frequency of occurrence' of the systems themselves. This is not a functional feature but Crystal includes it in this section rather than in the phonetically based section. What Crystal means by this is unclear. There will surely be a difference in frequency of occurrence between the two systems since the prosodic system will always be present in speech (there will always be some pitch, some duration, some loudness or some silence in connected speech), whereas the features which make up the paralinguistic system need not be present in speech. There must therefore be a difference in frequency of occurrence between the two systems. There is however confusion about the domain of the systems, on p. 137 Crystal says that "in practice (paralinguistic features) seem to be frequent only on single words or short sequences of words", yet when Crystal talks about identifying a particular paralinguistic feature (p. 136) he says:

"It is an effect which cannot be explained by reducing the phenomena to a perception of individually unrounded vowels .. in most cases the overall

effect is cumulative; a product of our perception of a sequence of syllables articulated in a basically identical way."

There seems to be a confusion therefore between the length of connected speech necessary to identify a paralinguistic feature, and the length of connected speech which is in fact identified as possessing a particular paralinguistic feature. Theoretically however, a paralinguistic feature may occur over any length of utterance so perhaps the above anomaly is unimportant.

Because of basic problems in distinguishing between these two systems Crystal notes (p. 138) that there were in fact occasions when the use of a vocal effect was ambiguous, due to a feature from the paralinguistic range formally coinciding or overlapping with a characteristic of an individual's voice quality. Voice quality is another suprasegmental system defined by Crystal as "the permanent non-segmental idiosyncratic factor in a person's speech." (p 126) Creak was the most frequently ambiguous feature, but Crystal says that the two functions of this feature should be identifiable since the feature of voice quality would "have a predictable distribution related to the occurrence of certain prosodic features of pitch range and tone (such as falling tone, low height) and be inexplicable in relation to situational motivation; whereas, if linguistic, the

feature will occur in particular lexical and situational contexts and produce a definable attitudinal reinforcement." (p. 138) This last quote may well give us the clue to the differences between the three suprasegmental systems mentioned above. Voice quality is a permanent feature of an individual's speech; the prosodic system is a system involving different systems of contrast which will always be present in connected speech; the paralinguistic system operates occasionally in speech as an emphatic marker of an attitude which is already present in the utterance (cf. p. 137 where Crystal says that the function of paralinguistic features "might well be to give additional emphasis or pointing to an attitude already present in an utterance, either in the prosodic features or in the lexis.") These three systems cannot be easily distinguished according to the phonetic parameters which signal the systems since the parameters are often easily confused if not exactly identical.

We might therefore propose an alteration of Crystal's views as follows: voice quality is the suprasegmental system of speech which is permanently present in connected speech and which serves to identify one individual from another. The prosodic system is the system of contrasts signalled by the parameters of pitch, loudness, duration and silence in their neutral range. The paralinguistic system employs the features

which make up both previous systems and uses these features to emphasise the attitude which is already present according to lexis or context of situation. The paralinguistic range, then, can be said to employ the above phonetic parameters in the extended range which operates around or beyond the neutral range reserved for the prosodic system and the 'normal' features of an individual's voice quality.

Let us now examine some of the systems which constitute the prosodic system as Crystal describes them. Within Crystal's cover term 'prosodic system' we have several discrete systems in operation, ranging from the 'most linguistic' systems of tone; pitch range; tempo and loudness, to the 'less linguistic' systems of rhythmicality; tension and pause. We will spend most of our time with the 'most linguistic' systems.

Pitch range: this system is defined by Crystal as follows (p. 143-4):

"For any speaker, the first prominent syllable of a tone unit is articulated at or around a stable pitch level for the majority of his tone units ... Once the pitch level of the onset is determined, the pitch range of the remainder of the tone unit is defined in relation to this by plotting the various upward or downward movements which take place."

Thus the pitch range of any unit is based on the relationship between the 'constant' first prominent syllable and the syllables following it. Yet when Crystal goes on to define his simple pitch range system he states that (ibid.) "contrasts exist in a system of six marked terms around a norm" for pre-nuclear stressed syllables, and his terms 'drop', 'low drop', 'continuance', 'booster', 'high booster', and 'extra-high booster' are then defined according to the relationship between any two syllables in the prehead, A and B. The only condition involved is that the syllable B must be stressed but A may be either stressed or unstressed. Therefore B could be the first stressed syllable of the utterance preceded by an unstressed syllable A. If the full system of contrasts can operate on the pitch range of the first stressed syllable then, how frequent is the occurrence of the 'norm' i.e. the "stable pitch level" on the first prominent syllable, and how is it determined?

Crystal states that the pitch range where the stressed syllable B is lower than the preceding syllable A (which can be stressed or unstressed) is the normal, unmarked term in the pitch range system. But if, as an instance, we take B as the first stressed syllable of the tone group, then none of Crystal's examples are examples of this 'norm' (cf. p. 145). Crystal's definition would only work if he defined the

two syllables A and B as the first two stressed syllables in the tone group such that A is the constant first stressed syllable and B is the varying second stressed syllable. The system of contrasts would then refer to the variable pitch range between A and B with A being 'normally' constant.

Loudness: Crystal chooses to differentiate between 'stress' and 'accent' as follows (p. 156); "loudness was defined as the auditory correlate of amplitude and stress as those variations in linguistically contrastive prominence primarily due to loudness; variations primarily due to pitch were called accent."

Crystal differentiates between these two parameters and sets up a pitch range system to identify contrasts signalled primarily by pitch, and an independent loudness system to identify contrasts signalled primarily by loudness, these two systems lean so heavily on each other that very few contrasts seem to be in fact independent.

The pitch range system is divided into two sub-systems -- one operating on stressed syllables and the other on unstressed syllables, therefore the notion of stress must be 'given' independently in definable terms before we can use the pitch range system. But, conversely, the notion of 'marked' versus 'unmarked' in the pitch range system is held to be 'given' in the stress system.

Crystal identifies six terms in his loudness system. Category E (strong or extra stress) can be identified independently in terms of "a marked increase in loudness" (except that we are never told what an unmarked increase in loudness is) which can occur on any syllable. Four of the remaining five terms depend, in their definition, on the co-occurrence of an 'increase in loudness' with a marked syllable in the pitch range system. Since there is no way of defining 'an increase in loudness' without a norm to refer to (which we are never given) we must turn to the term 'marked' in the pitch range system. An unmarked syllable in the pitch range system is a stressed syllable which is slightly lower than any previous syllable, therefore any other formation must be regarded as marked. According to the above definitions, then, it seems to be the case that the only way we can identify a stressed syllable is when we already have a stressed syllable which is not lower than the preceding syllable, i.e. we can't! Crystal's sixth term in the loudness system is the category 'unstressed' which refers to "all remaining syllables regardless of pitch height" (p. 159)

The above two systems would seem to be highly inter-dependent to say the least, in fact Crystal notes on the same page (p. 159) "the experimentally proved correlation of increased stress with higher pitch". Crystal is only mentioning this in connection with strong

stress but this notion should obviously be extended such that the two systems are amalgamated into one, with each corroborating the other. It has been shown time and time again in experimental work on the acoustic correlates of stress that pitch and loudness (or rather their acoustic correlates fundamental frequency and amplitude) are intimately connected (see chapter 1 and 2 for more details), therefore it would be a very natural thing to do to subsume the two parameters under one system. Crystal gives no satisfactory reasons for separating the two.

Tone: Crystal's system of tone is discussed in the context of the tone unit:

"Minimally, a tone unit must consist of a syllable, and this syllable must carry a glide of a particular kind. This is the obligatory element, and it is usually referred to as the nucleus of the tone unit ... Maximally, the tone unit may consist of three other places, or segments, though only the first two of these are normally independently contrastive: the head, the pre-head, and the tail."
(p. 207)



This 'obligatory element' is also referred to as the nuclear tone, and it is this syllable ("the most prominent syllable of a tone unit" p. 142) which car-

ries the contrastive system of tone. Although this syllable can be identified in functional terms as the only obligatory element of a tone unit, it should be noted in passing that it may not be so easy to identify such an element in phonetic terms. Crystal says above that it is the "most prominent syllable", yet on p. 120, in a footnote he says "Prominence has ... no theoretical status in this framework, being used to refer to the general distinctiveness of an utterance of any length."

Let us firstly examine the system of tone, and later on we will discuss the tone unit at more length.

Crystal divides ~~the~~ system of nuclear tone into three sub-systems: simple, complex, and compound. The simple tone system involves three tones with a uni-directional pitch movement: rising, falling, and level. The pitch height is not taken into consideration within this system, this is covered by the pitch range system as is the amount of movement. The pitch range system therefore is called into play for a value in the high / low system and in the wide / narrow system. It is interesting to note that Crystal considers it quite sufficient to mark the beginning point of the pitch in terms of height. He excludes the necessity of accounting for the pitch of the end point by defining the amount of movement or width of pitch movement. He selects the beginning point rather than the end point "partly because of the formal indeterminacy which exists at the

end of the glide"... (and also because the end-point of a glide) "does not have any explanatory value elsewhere." (p. 213-4) Yet on the same page as the above quote, Crystal talks about the "high-mid falling tone", where he specifies both the beginning and the end points of the movement. There may well be contrasts expounded by the end point rather than the beginning point in different varieties of English, for example the often noted 'fall to mid' in Scots. Such contrasts would not be easily expressed by Crystal's present system.

Crystal's second sub-system is the complex tone system. This system involves changes of direction in pitch movement within a syllable with only one maximum of prominence. The terms included in this system are the fall-rise, the rise-fall, the rise-fall-rise, and the fall-rise-fall. Crystal notes that the first part of the first two terms mentioned (i.e. the fall-rise and the rise-fall) is phonetically the most prominent yet he says the unmarked phonetic form of these terms is generally agreed to be  and  which looks as though the fall is the most prominent in both cases.

Crystal's third sub-system is the compound tone system. This system refers to combinations of two kinetic elements of different major phonetic types acting as a single tonal unit. The tones in this system must be endocentric and are of the two types rise-fall, and fall-rise. There must be no evidence of a tone unit

boundary but rather a continuous movement between the two tones. One element must be more phonetically prominent than the other (normally the first) whereas the second element is the major functional element and forms the basis for the classification of the tone.

There are clearly several ways in which the latter two systems can be differentiated between, but I think the major difference between the two systems has to be that in the complex tone system there is only one point of prominence whereas in the compound tone system there are two points of prominence.

In Crystal's book The English Tone of Voice (1975), rather than forming systems based on the phonetic forms of the elements involved, he relates the tones to their grammatical function saying (p. 36):

"By distinguishing the three phonetic variables which underlie the above nuclei (viz. general pitch direction (falling-type v. rising-type), pitch-range (high v. low start) and complexity (viz. \wedge , \vee)), the following systematic relationship might be hypothesised between tonal type and grammatical function:"

\	NEUTRAL	———	+↑	= CONTRASTIVE FOCUS
/	CONTINUITY	<div style="display: inline-block; width: 50px; border-left: 1px solid black; height: 10px; margin-left: 5px;"></div>	+↑	= CONTRASTIVE QUESTION
		<div style="display: inline-block; width: 50px; border-left: 1px solid black; height: 10px; margin-left: 5px;"></div>	+complex	=NEG. IMPLICATION

The grammatical functions of the tones described are then related to the semantic functions of the tones presented by Crystal as follows (ibid., p. 39):

final: neutral

↓ \	non- personal definiteness	$\left\{ \begin{array}{l} + \uparrow \text{definite emotional} \\ \text{commitment} \\ + \text{complex: definite} \\ \text{outcome} \end{array} \right.$
↓ \	final: unsociability	

— \	final: absence of emotional involvement
— \	non-

final: implication of routineness

final: personal inconclusiveness
social openness

↓ /	$\left\{ \begin{array}{l} + \uparrow \text{definite emotional} \\ \text{enquiry} \\ + \text{complex: uncertain} \\ \text{outcome} \end{array} \right.$
↓ /	

non-

final: neutral

Crystal goes on to say:

"Grammatically, all that can be said about complex tones is that they are distributionally restricted to the ranges of the simple tones of the same phonetic type: for example, \wedge may be used in all plac-

es where $\downarrow \backslash$ or $\uparrow \backslash$ go, and simply adds attitudinal information to the utterance. If \wedge were to replace, say, $\downarrow /$, however, then there would be a corresponding change in grammatical function, but this would result, not from the pitch-complexity of the tone, but from its status as being basically falling in type."

Thus, in Crystal's original proposals as outlined in 1969, the nuclear tones are sub-classified according to the phonetic parameter of pitch movement and its relative complexity, but in his later model as presented in 1975, Crystal sub-classifies the nuclear tones into basic direction-types of pitch movement, relating each type to a grammatical function and an attitudinal core meaning. A more complex nuclear tone is then said to add attitudinal intensification, or attitudinal information. These two models, therefore, present two quite different ways of approaching the problem of sub-classification of nuclear tones.

Let us return to the discussion of the tone units themselves. Crystal describes the identification of tone units as follows:

"In English there seem to be regular definable phonological boundaries for tone units in normal (here meaning mainly 'not too hurried') speech.

Given that each tone-unit will have one peak of prominence in the form of a nuclear pitch movement, then it is the case that after this nuclear tone there will be a tone-unit boundary which is indicated by two phonetic features. Firstly there will be a perceivable pitch-change, either stepping up or stepping down, depending on the direction of nuclear tone movement -- if falling, then step-up; if rising, then step-down; if level, either, depending on its relative height... The second criterion is the presence of junctural features at the end of every tone unit. This usually takes the form of a very slight pause, but there are frequently accompanying segmental phonetic modifications (variations in length, aspiration, etc.)."

(1969, p. 205)

The tone unit itself, as noted previously, must minimally consist of a nuclear tone, and maximally be of the form:

Prehead	Head	Nucleus	Tail
---------	------	---------	------

The tail is normally non-contrastive in English and is usually determined by the nucleus, although (p. 223) "occasionally linguistically significant variation may occur". The remaining systems of prehead and head

are said to operate contrastively with no obvious inter-relationship between them, though each depends to a certain extent on the pitch direction of the nucleus.

On several occasions I have shown in this chapter that Crystal's overall tendency to prefer several systems operating independently from each other tends to obscure generalisation and results in a considerable amount of overlap. To propose 4 semi-independent systems operating within the tone unit then, may well be another example of splitting up an autonomous 'whole' into its component parts. The next chapters will examine some of the main points summarised in this chapter in the light of data base collected for the SSRC project.

Footnote to chapter 3:

1. The table as presented above does not capture generalisations such as the one Halliday mentions concerning tone 4 which is said to function as the neutral tone where a conditional clause precedes a conditioned clause. But since such a context was not mentioned when setting up the mood system the above table is restricted to the elements of the mood system and Halliday's tone system.

Chapter 4.

4.1.0 Data collection

Most of the previous work on intonation has been based on data consisting of isolated sentences or short texts read aloud. Recently, Crystal (1969, 1975) has criticised previous work involving such a data base as 'extremely restricted', and he has claimed that the analysis of 'quantities of spontaneously produced material' is the only valid approach to a study of intonation. Crystal's criticism of studies which have exclusively used forms of reading is certainly sensible but his total rejection of spoken prose as a source is unwise I think, and his exclusive use of connected spontaneous speech as a data base is a misguided overreaction. Both spoken prose and spontaneous speech should be included in the data of any intonation study so that the analysis of one data type complements and contributes to the analysis of the other in a valuable way. Spoken prose allows the investigator to compare the same lexical content spoken by different speakers without the false starts, hesitation pauses, incomplete sentences and other performance variables which are inevitably involved in spontaneous speech. The data sample used as a basis for this thesis, then, included a read text, short responses to a questionnaire, controlled game situations, spontaneous speech based upon a particular photograph and informal conversation of various kinds.

These materials are now described in detail.

4.1.1 Text

The Witch's Daughter

1 In the midst of a range of wild mountains was
2 a small straw hut, where an old man lived with his
3 three sons. Every day the father went out to look
4 for fuel.

5 Once he met in the wood an aged widow in white
6 clothes, who was seated on a square stone playing
7 chess. Since the old man was a keen player himself,
8 he asked the woman, "May I watch the game?" She re-
9 plied, "Do you want to play with me?" "Certainly",
10 said the old man. When the widow asked for what
11 stakes they should play, he suggested playing for
12 his wood. But the old woman said, "No, we can't
13 play for wood, because I don't have any wood. How
14 many children have you, though?"

15 "Three sons", was his answer. "Three sons? That is
16 perfect. I have three daughters. If you win, I will
17 send them as brides for your three sons; but if I
18 win, you must send me you sons as sons-in-law". The
19 old man stroked his beard for a while, but finally
20 gave his assent.

21 He lost each of the games they played, and
22 when the widow got up to leave she said, pointing in-
23 to a dark valley, "There is my house. Tomorrow send

24 me your eldest son, three days later the second,
 25 and again after three days the youngest." She then
 26 departed, and the old man went home, without collect-
 27 ing any more wood, to tell his sons what had happen-
 28 ed. How pleased they were when they heard it!

29 The next day he sent the eldest son; three
 30 days later, the second; and on the sixth day he
 31 sent the youngest.

Each informant was asked to read aloud the above text. It was left until such time as the interviewer felt the informant relaxed and used to the interview situation before the informant was asked to read the text. The informants were members of various age groups, social classes, educational backgrounds, occupations etc, therefore the text selected was short, included fairly simple vocabulary, and its subject matter was not esoteric, involved or difficult.

Note that the text is not an incomplete excerpt, but has a beginning, middle, and an end and is divided into four orthographic paragraphs. These features were motivated by an interest in the reader's treatment of orthographic paragraphs when reading aloud, such that the notion of 'key' -- (the use of the speaker's tessitura to divide up a text into parts) -- and the concept of 'paratones' -- (intonation units which have a larger domain than the tone-group) could be examined

with relation to the read text. (cf. chapters 6 & 7)

In addition to the above requirements it was decided that the following clause types should be included (numbers to the right refer to the line/s in which the clausal type occurs in the text.)

declarative	(3,5,16 and others)
exclamation	(28)
imperative	(23)
interrogative	
inverted	(8)
do-support	(9)
wh-question	(14)
echo-question	(15)
conditional	(16-17)

also:

conjoined clauses of equal status	(16-18)
direct speech	(8,9,12,14,15)
reporting phrase following direct speech	(10)
reporting phrase preceding direct speech	(8)

4.1.2 Sentences read aloud

After eight interviews had been carried out it was decided to add a set of sentences to the material to be read aloud. Certain clause types in the text were

emerging as being very interesting, and it was felt that a larger corpus of these particular clause types would be useful to have. The sentences are as follows:

1. They missed the last bus so someone suggested a taxi, but they didn't have enough money so they walked.

(sentence 1 resembles the sentence beginning on line 25 of the text. Both have a succession of clauses).

2. Since the old man was a little deaf he always asked people to speak up.

(sentence 2 and sentence 13 - see below - form a contrasting pair. We wanted to observe the intonation of a phrase such as "he always asked people" before direct speech and before indirect speech.)

3. How many children do you have?

(sentence 3 contrasts slightly in form with "How many children have you, though?" lines 13-4).

4. The old man asked if she had posted the letter but the nurse said "No".

(sentence 4 contrasts with sentence 8 - see below, and resembles the sentence beginning on line 12: But the old woman said...)

5. He played squash on Mondays, Wednesdays and

5. ... Fridays.

(resembles final sentence of text but the elements in the sequence are single words rather than short clauses).

6. Andrew wanted to phone his mother so he asked a passer-by "Have you change of a £ 1?"

("Have you change of a £ 1?" is standard Scots. Sentence 6 contrasts with 11 and 14 below. We wanted to be able to compare a reporting phrase preceded by a clause, as in 6, and the same reporting phrase not preceded by a clause, as in 11.)

7. They missed the last bus so someone suggested taking a taxi.

(see 1. This sentence was included to contrast with 1, enabling us to compare the intonations of the clause 'so someone suggested taking a taxi' when it is in final position and when it is in non-final position.)

8. The old man asked if she had posted the letter.

(see 4 above).

9. Every day he did the Scotsman crossword.

(resembles second sentence of text, - was included to enlarge corpus of sentences beginning with a time adverbial).

10. On Monday she got the gas bill, three days lat-

10. ... er she went to the bank and on Friday she paid it.

(resembles final sentence of text)

11. He asked a passer-by, "Have you change of £1?"
(see 6 above)

12. "Would you like a Vodka Gimlet?" "A Vodka Gimlet?"

(this sentence was included to enlarge the corpus of echo-question read aloud.)

13. Have you change of a £1?

(see 6 and 11. We wanted an example of direct speech not preceded by a reporting phrase or other contextualising phrase).

4.1.3 Questionnaire

It was felt that the sample should include spontaneous speech that was not long and rambling with a lot of hesitation pauses and replanning, that is, short stretches of spontaneous speech were required. The best means of eliciting this type of spontaneous speech was to ask informants to answer simple questions such as "Where were you born? What is you address?" The questions were not puzzling or thought-provoking but were about everyday concerns and general facts because we wanted to elicit short, automatic responses. This 'question-answer' section had a dual purpose: like the text, it would make analysis of speech into tone groups, identi-

fication of tonic syllables, etc. easier (short stretches of fluid speech are much easier to analyse in these terms than long involved stretches); and it would provide vital information about the background of the informants which might be relevant to the analysis. The following questions were on the questionnaire. Reasons for including each question, when not obvious, are noted in parenthesis.

1. Could you give me your full name?
2. Could you spell your first / middle / last name?
(we thought spelling would be interesting from the point of view of grouping and sequencing, i.e.
Are the same strategies used when the elements are of one syllable and when they are more complex?)
3. Miss or Mrs?
4. Could you give me your address? (needed for future contacting of informants)
5. Where were you born? (linguistic background)
(if Edinburgh) Have you lived in Edinburgh all your life?
(if yes) Have you moved around at all in Edinburgh?
(if not born in Edinburgh) Where else have you lived? How long have you lived in Edinburgh?
6. Where were your parents born? (linguistic background)
7. Where did you go to school? (educational background)
8. (as appropriate) Do you have any (brothers and

8. ... sisters?)
(children)
9. Could you give me their names and ages? (possible other informants; to elicit a list of elements)
10. Are they at home / school / college? (to elicit a list of elements which are short)
11. Could you describe the route you take to get from your home to your school / place of work? (to elicit a list of elements which are long.)

4.1.4 Echo-question elicitation

Because of suspicion that the intonation of echo-questions in Scottish English is different in form from echo questions in Southern educated English, it was decided to include in the interview a device to elicit an echo question spontaneously. The following means was devised. The informants would be shown a photograph and asked to identify where it was taken. (This was their task as far as they knew). The photograph would contain an object which was likely to be unfamiliar to the informant or which had a strange-sounding name. After asking the informant to identify the place shown in the photograph, the interviewer would then ask (ostensibly as a helping hint) "Do you recognise the (odd name)?" the implication seeming to the informant to be that if he recognised this particular object or landmark he would then know where the photograph was

taken. The actual purpose of the question was to get the informant to query the constituent in the 'hint-question' using an echo-question. A photograph of Brown's Square in Edinburgh in the 1850's was selected showing a part of the city which has since been drastically renovated and would therefore be difficult to identify. The landmark with an unfamiliar name was the tower of St. Giles Cathedral, which is called a 'lantern' because of its shape. This terminology is not known to many people. Informants were shown the photograph, asked if they recognised it, and then given the hint "Do you recognise the lantern?" This method successfully elicited an echo question more than 50 per cent of the time, not exclusively on the constituent lantern, but occasionally on some other name which came up during the discussion of the photograph. Out of 30 attempts 11 echo questions were received on the constituent the lantern and 7 echo questions on other constituents. It was found that the best responses to the photograph came from the over forty-five age group who had some experience of and an interest in how Edinburgh looked in the past. From the 15-25 age group the response was poor, so another more recent photograph of Edinburgh was added to the interview, which, though not as successful in eliciting an echo question, did stimulate spontaneous conversation.

4.1.5 Spontaneous speech

In addition to the speech sample collected from reading aloud, short responses to questions, and discussions based on the photograph, free spontaneous conversation from each informant was also required. To obtain this, the tape-recorder was put on before the interview proper began, so as to collect various speech behaviour such as introductions, directions and responses to these directions. (Could you sit nearer the microphone? Yes, of course), queries by the informants as to what was to be done, how long it would take, etc.; and then left on after the interview was over to record comments, leave-taking, etc. During the interview proper the interviewer would attempt to lead the informant into casual relaxed conversation using whatever taking-off-point seem appropriate (the photograph, content of the text / sentences, answers given to the questions in the questionnaire, fear of reading aloud, etc.). From some informants very little spontaneous conversation was collected, from others a great deal. A range of (spontaneous) speech styles, such as narrative (relating an experience), dialogue (small talk, exchange of information /opinions, banter) were collected from this section of the interview. A recording was also made of a longer conversation between 3 informants at a political and social club who did not realise that the tape recorder was recording. The spontaneous speech data is,

therefore, not restricted to the interviewer / informant dialogue.

4.1.6 Game situation

In order to obtain samples of one specific speech style, a game situation was set up with two players. One player (P1) was given the text of a story to read and become familiar with (see p.241); the second player (P2) was given a list of actions or attributes and a list of characters (see p.244). The object of the game was then for P2 to reconstruct the story by matching the characters and actions, and then ordering the events. P2 did this by asking questions which P1 was allowed to answer by saying 'yes' or 'no'. By restricting the information which P1 was allowed to give, we hoped to obtain specific types of questions from P2 such as cleft questions -- 'Was it the rich farmer who had three sons', and contrastive questions or statements such as --

P2: Did the old man run away?

P1: No

P2: Well did the rich farmer run away then (where
the underlined element
is contrasted)

Many samples of the above structures were in fact

obtained. The game itself proved to be sufficiently complex for the players to concentrate only on the task at hand, being unaware during the recording that the speech style was being investigated rather than the strategies involved in the game. We can therefore say that the recordings of these games were samples of one specific style of spontaneous speech.

All of the above types of speech formed the data base for the SSRC Project called the Intonation of Scottish English. The data base for this thesis was drawn from the project data base which was made available to me.

4.2 Analysis of the data

The data base was always analysed auditorily by myself and one other person simultaneously in order to try to reduce the 'subjectivity' of an auditory analysis. My principal co-analysts were Joanne Kenworthy (an American English speaker) and Roger S. Brown (a speaker of Southern British English) to whom I am deeply indebted. Neither of them were speakers of Scottish English, therefore there was always one 'native' speaker (myself) and one 'non-native' speaker collaborating on the auditory analysis. This meant that the different parameters which were brought to bear on the analysis had to be resolved before final agreement about the transcription was reached, i.e. the 'native' speaker might be familiar

with certain patterns and expect to hear them, but if these patterns were unfamiliar to the 'non-native' speakers then a more objective analysis might be expected from the collaborators which would accompany the native 'insights'. When final agreement was reached the resultant 'impressionistic' transcription was marked down on a three line stave -- the bottom line representing 'low' in the speaker's range, the middle line representing 'mid' ⁱⁿ the speaker's range, and the top line representing 'high'. This transcription could then be converted into a 5-number system with the correspondence as follows:

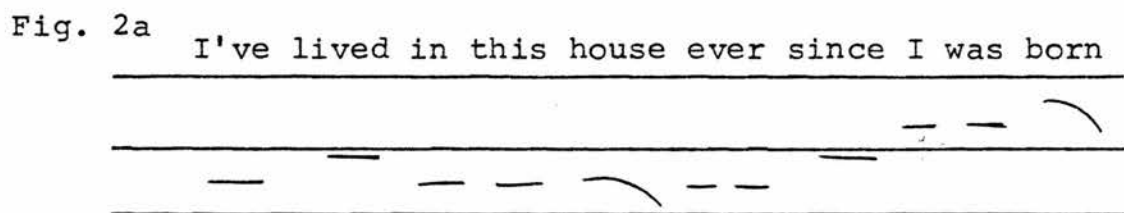
Fig. 1.

'high'	_____	5
	4	
'mid'	_____	3
	2	
'low'	_____	1

such that low=1; lowmid=2; mid=3; mid-high=4; high=5. (for further discussion concerning the use of a 3-line stave see R. S. Brown "Paratones: Their reality and realisation" in the SSRC Report, HR3601/1, October 1978).

It was interesting to note that as transcriptions progressed, these numerical values seemed to become highly correlatable with 'phonological' pitch heights, such

that level 1 was used as a marker of 'finality' occurring for example at the end of sentences rather than at the end of clauses; level 2 was used as a 'normal' base-line from which peaks of prominence departed; level 3 was used either as a 'lowered' or 'reduced' peak of prominence, e.g. preceding or following a level 1; or as a 'raised' base-line occurring for example in questions or exclamations; level 4 was used for a normal peak of prominence, and level 5 as a 'marked' or extra high peak of prominence occurring for example as the initial peak of a paratone (cf. chapter 7). It should also be noted that the analysts using the impressionistic auditory analysis rarely used the extremities when marking the transcription (visually) on a stave, but when analysing speech using the levels 1-5, the extremities 1 and 5 were used. Thus the correlation as stated in Fig. 1 should not be seen to be an exact correlation but rather an approximate correlation such that a stave transcription as follows (Fig. 2a) could be correlated with the numerical transcription shown in Fig. 2b.



Since there have been frequent arguments and confusion in the past about the number of pitch levels and the correlation between the level itself and the value of the numerical representation (i.e. whether low=1 or 5), throughout the thesis, examples of the data will be given with the 'amalgamated' impressionistic transcription on the 3-line stave underneath the orthographic transcription. The reference number of the informants will be given to the right of the example, together with the page and line number of the transcription such that 13/1/6 would refer to the informant number 13, page 1, line 6.

4.3 Initial examination of the data

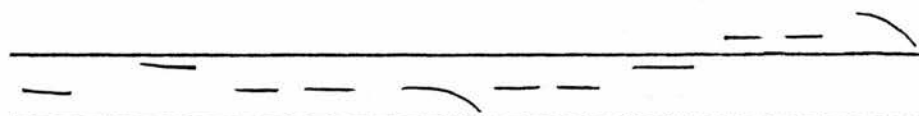
Since Crystal advocated the study of spontaneous speech so strongly, I first examined stretches of free conversation and attempted to analyse the examples into tone groups according to definitions and examples given by Halliday, Crystal and others. I encountered several difficulties, however. In Prosodic Systems and Intonation in English (1969), Crystal says there are regular definable phonological boundaries for tone units (or tone groups) in normal (not too hurried) speech. These boundaries are signalled by:

- (a) step-up (if the nuclear movement is falling)
- (b) step-down (if the nuclear movement is rising)

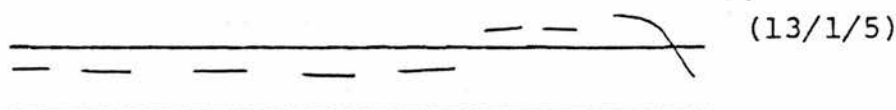
- (c) either (if the nuclear pitch movement is level)
depending on relative height.

Crystal states that these pitch changes are due to the fact that the onset of each tone unit in a speaker's utterance is at more or less the same pitch level. He also states that the junctural features at the end of every tone unit take the form of a very slight pause which is shorter than the minimal term in his pause system. There may also be segmental phonetic modifications (variations in length, aspiration etc.). According to Crystal "these phonological criteria suffice to indicate unambiguously where a tone-unit boundary should go in connected speech in the vast majority of cases". He admits that ambiguous cases can be 'thought up' as he puts it -- and then we must have recourse to grammatical or semantic criteria. However, the above criteria were not always sufficient for resolving doubts such as the following. Consider the following example:

1. I've lived in this house ever since I was born



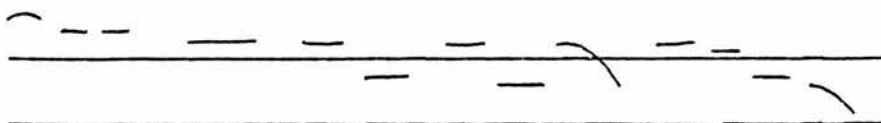
in fact that's for well sixty years



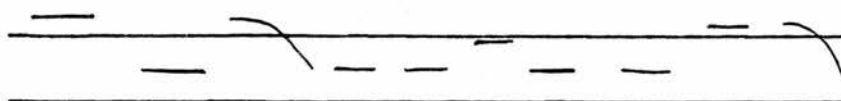
The terms tone unit and tone group will be used interchangeably for the present. The main difficulty in resolving this particular example into tone groups revolves around the phrase 'in fact'. Is 'in fact' a separate tone unit? Auditory and spectrographic analysis show no pauses on either side of 'in fact', and no perceivable 'step-ups' or 'step-downs' following it. If the phrase is not a separate tone unit then where does it belong? Does it belong to the preceding tone unit (i.e. I've lived here ever since I was born in fact) or with the following tone unit (i.e. in fact that's for well sixty years) or is it perhaps a separate tone unit? Can this decision be resolved by referring to grammatical or semantic criteria? I think not. Grammatical and / or semantic criteria could be used to argue for any of the above proposals. If 'in fact' is regarded as a lexical filler, this raises the question of lexical fillers in general -- should they be attached to a tone unit, or can they comprise a tone-unit by themselves? (see section 4.5 for further discussion)

The following example is another case where a problem phrase could belong to the tone unit on either side of it, but I don't think the phrase in this case could be classed as a lexical filler.

2. I only stopped over + overnight literally +



there last year I flew in from New Zealand



(23/7/1-2)

(+ represents a short pause)

This example has many problems which are not resolved by the criteria given by Crystal above. The problem phrase similar to the one in the first example is last year. Does it belong to the first phrase giving, I only stopped overnight there last year, or does it belong to the second phrase, giving, last year I flew in from New Zealand? Many people working on the intonation of sentences read aloud might predict that under circumstances such as these the intonation contour should disambiguate the sentence. If this is the case then the tone unit boundary should be easy to identify. In fact, when we try to identify the tone unit boundary according to specific phonetic criteria such as those proposed by Crystal, we find that the tone unit boundaries themselves are ambiguous.

If we mark the tone unit boundaries according to pauses in this sample of speech, the tone units would be as follows: (each unit being surrounded by brackets)

(I only stopped over) (overnight literally) (there last year I flew in from N.Z.)

There are three tone units according to pause criteria, but the last unit looks as though it contains more than one semantic or grammatical unit. If we define the tone unit boundaries according to grammatical / semantic criteria we might propose tone units such as the following:

(I only stopped over-overnight (literally) there last year) (I flew in from N.Z.)

In other words we have one major syntactic break between year and I plus an aside / parenthesis which might be considered as metalanguage in the middle of the first tone group literally. These two analyses then, give us different results for the placement of the phrase last year. According to the pauses, this phrase belongs with the last tone unit, but according to semantic / grammatical criteria the phrase last year should perhaps belong to the first major unit since it would be much more likely for this phrase to be attached to the end of a unit than to prepose it to the front of a unit when there is no obvious reason for doing so, i.e. there is no contrast involved. Thus on semantic grounds the phrase belongs to the first unit, but on phonetic grounds the phrase belongs with the last tone unit. There seems to be an ambiguity here which cannot be resolved.

Let us therefore turn to Crystal's other main phonetic criterion to see whether this would resolve the conflict. Crystal states that there is a tone unit boundary when there is a step up (if the nuclear movement is falling) or a step down (if the nuclear movement is rising). Let us therefore divide this sample of speech into tone units according to the latter criterion. The tone units would be as follows:

(I only stopped over) (over) (night) (literally)
 (there last) (year I flew) (in from N.Z.)

The fundamental problem when dealing with tone unit identification according to the criteria of step-up / down, is the problem of identifying the nuclear movement in the first place. We have to be able to identify the nuclear pitch movement before we can identify the step-up / down which marks the tone unit boundary, and if the nuclear pitch movement is defined in terms of 'the maximum pitch movement in the tone unit' we have come full circle with no way of escaping from the circle.

Perhaps if we match up these several means of tone unit identification we might define the boundaries of the tone unit according to the criterion 'the more often a boundary is defined, the stronger (or more plausible) it becomes'. In other words if several different criter-

ia are set up as criteria upon which the identification of the tone unit is to be based, we might propose that we require a certain percentage of these criteria to co-occur in order for a tone unit boundary to be set up. If the boundary is only defined according to one criterion out of say four, this might not be considered a strong enough reason for positing a tone unit. Alternatively the criteria might be ranked in some way such that pause criteria outranked pitch criteria etc. Let us compare the criteria which have been proposed so far and the tone units which they define in order to see where the strongest boundaries occur. The boundaries will be defined as follows: under a) according to step-up/down; under b) according to pause; under c) according to grammatical or semantic phenomena.

a) according to step-up / down

(I only stopped over) (over) (night) (literally) (there last) (year I flew) (in from N.Z.)

b) according to pause

(I only stopped over) (overnight literally) (there last year I flew in from N.Z.)

c) according to grammatical or semantic criteria

(I only stopped over- overnight (literally) there last year) (I flew in from N.Z.)

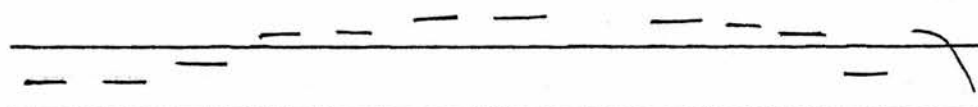
If we compare the above three versions of possible tone unit boundaries for the above speech sample we find that the only boundary marked by all three criteria is the final boundary after New Zealand. The two phonetic criteria then agree that there should be boundaries after over, literally and New Zealand but none of these boundaries is supported by the semantic or grammatical criteria.

This is a basic dilemma for Crystal's analysis. He wants to define tone unit boundaries according to phonetic criteria alone, but finds that this cannot be done one hundred per cent of the time. He therefore allows himself to have recourse to grammatical and semantic criteria when in difficulty. In doing this, however, he lays himself open to the problems which arise when there is conflict between the phonetic criteria and the semantic / grammatical criteria. Which one takes precedence? If we are defining a phonological unit then perhaps the phonetic criteria should take precedence, but if we are defining an information unit (e.g. see Halliday 1967) or a neurological unit of some kind (see evidence from slips of the tongue in Boomer and Laver, 1968; also Goldman-Eisler 1964), then perhaps the semantic and/or the grammatical criteria should take precedence. (For further discussion see chapter 5).

4.4 Pause phenomena in the data

One area where Crystal's definition of tone unit boundaries creates problems, especially in spontaneous speech, involves the phenomenon of the planning pause. It is a well known fact that people, when speaking spontaneously, frequently pause while they search for a lexical item or plan ahead or re-plan an utterance (cf. Boomer 1965, Goldman-Eisler, 1967, etc). We might refer to this phenomenon as a planning pause. If Crystal is going to define tone units as being surrounded on either side by a 'very slight pause' then problems arise when we have to decide whether a pause is defining a tone unit boundary, or when it is functioning as a hesitation or planning marker. The following sample is an example of this problem in the data:

3. Was the bad-natured daughter ++ beautiful as well



A+D/1/5

In the above example there is a fairly long pause but there is no pitch movement on either side of the pause, rather the pitch movement is 'rounded off' after the pause. The unit rises up to level 4 on daughter, continues at the same level after the pause, then falls to level 1 at the end of the unit. Therefore the pitch

movement would predict that this sample of speech is a single unit (although there is some movement on well, but for further discussion of this see chapter 5).

Crystal does not really help us to identify the various forms of pause nor their functions. He says (p. 206) that the pause which delimits the tone unit is 'a very slight pause' ... which 'is shorter than the minimal term in the pause system'. But he never defines the length of pause in the pause system except in terms of the interval of an individual's rhythm cycle (p.171). Thus a unit pause would be 'the interval of an individual's rhythm cycle from one prominent syllable to the next'. The shortest term in the pause system is the brief pause which is 'approximately half as long as unit length'. According to Crystal's definition therefore, the pause which surrounds the tone units should be less than half as long as unit length. Conversely, there is a passage later in the same book (Crystal 1969) where he discusses the relationships between tone units with reference to pauses, and he has this to say (p. 243):

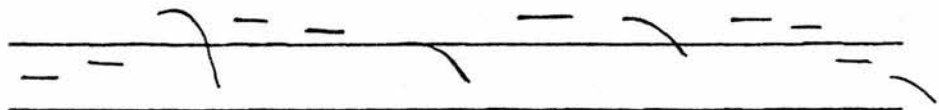
"... only unit pause displays sufficient frequency of occurrence at tone-unit boundaries to delimit sequences practically. (There are too few instances of double pause or more to produce a reasonable textual division; and brief pauses are altogether

too frequent, occurring as often within tone units as between them)".

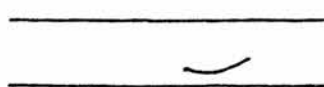
We must deduce from this that not only do brief pauses occur frequently at tone unit boundaries and within tone units themselves, but also that unit pauses and double pauses can occur at tone unit boundaries, so how is one to tell when a pause is occurring in the middle of a tone group or at either end? It is obviously not the case that the only pause which delimits a tone group is one which is "shorter than the minimal term in the pause system" if brief pauses, unit pauses and double pauses also occur at tone unit boundaries.

Pauses seem to be a multi-functional phenomenon. We have noted above that pauses occur when the speaker is planning ahead, or replanning an utterance. This was referred to as a planning pause. The planning pause is easily identified since it interrupts constituents such as the noun phrase, or the prepositional phrase. Note the following examples:

4. and my son he went tae + Gracemount Secondary



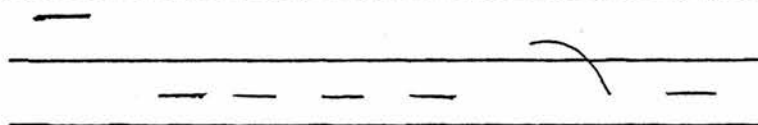
(mhm) + School +



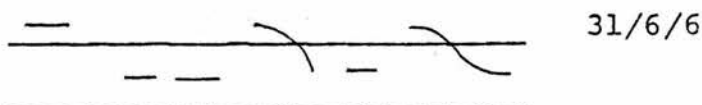
31/4/4

Example 4 shows a planning pause interrupting a prepositional phrase (tae Gracemount Secondary) plus a second planning pause interrupting the noun phrase (Gracemount Secondary School). Example 4, therefore shows the planning pause in two different environments.

5. that's the way that they bribed + eh +



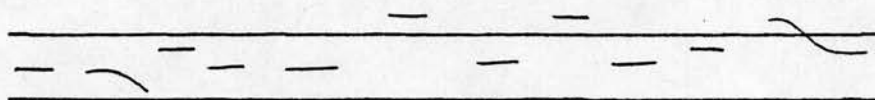
bricklayers too you know +



Example 5 shows a hesitation pause occurring between a verb and its direct object. This pause can also be referred to as a 'filled pause'. A filled pause will normally occur in some sort of planning or replanning situation but instead of there being a period of silence the speaker vocalises during the pause. In Scots, this vocalisation tends to be in the region of cardinal vowel 2 or 3, whereas for RP speakers this vocalisation tends to be in the form of a central vowel. A re-planning or planning pause may also be filled by 'lexical fillers' but these will be dealt with later (see section 4.5).

Another example of the planning pause is shown in no. 6.

6. and eh many + eh + sub-let houses you know +



(10/6/5)

Example 6 shows a filled pause occurring with silence surrounding it (as in example 5, i.e. many + eh + sub-let, and also a filled pause occurring without silence surrounding it as in and eh many. Thus there are different realisations of the term filled pause. The filled pause can occur with perceptible silence on either side, or it can occur without silence on either side.

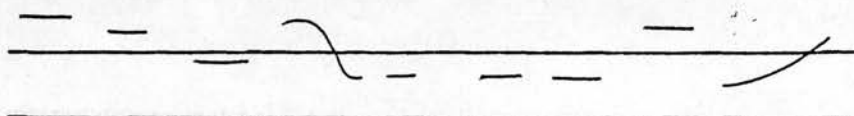
Although it seems relatively easy to recognise planning pauses (filled or otherwise), because pause is multi-functional we must expect there to be a confusion of functions at times. The following two examples would seem to be further examples of planning pauses.

7. oh its classed as a + slum a slum area now +



(31/8/6)

8. that is St. Giles eh + in the background +



(10/2/1)

Let us first examine example 7. The pause occurs between an article and its head noun therefore we seem to have a clear case of an example of a planning pause. The head noun is then repeated but we find that it has become a qualifier preceding another head noun area. We might refer to this phenomenon as an example of re-planning, but we could also refer to it as an example of repetition or even a false start. All of these things are involved (repetition, since slum is repeated; false start, since the status of slum changes from head noun to qualifier; and pause, since there is a period of silence) so perhaps we should regard the term planning as a cover term which can be realised by all of the above parameters. There would therefore be a confusion of parameters in example 7, but they would all be subsumed under one function, i.e. planning (see section 4.5).

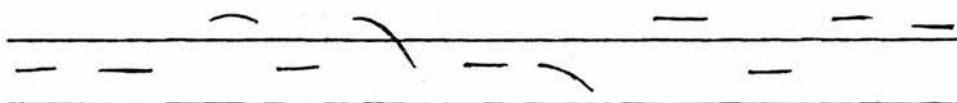
Example 8 involves a slightly different problem. There seems to be a planning pause here since we have a lexical filler followed by a perceptible period of silence, yet if we examine the pitch movement, the pause seems to divide the speech sample into two, such that each stretch of speech contains one large amount of pitch movement. Thus the pause in this case may well be functioning as the indicator of a tone unit boundary. Since there is a vocalisation immediately preceding the pause with the characteristics of a filled pause we might

propose that the two functions coincided in this example, i.e. the speaker chose to pause in order to plan ahead, at a tone unit boundary. Thus the pause in example 8 may well indicate that the speaker is replanning or planning ahead, but he is doing this planning at a tone unit boundary. This would seem to be a reasonable hypothesis since speech has to be chunked in some way in order to allow the Hearer and the Speaker to process speech in units greater than the 'word'. The tone unit would seem to be a logical neurological 'planning' unit. (For further discussion in connection with this idea see Goldman-Eisler, 1964 and 1967; Boomer and Dittmann, 1962). If the pause in example 8 is being used to define the boundaries of the tone unit, as well as functioning as a 'planning' pause then these two different functions should be noted and be treated separately. If a pause delimits a unit of some kind which is correlatable with a semantic or grammatical unit, then this is a boundary pause determining the boundaries of such a unit. If the pause occurs in the middle of a grammatical or semantic unit, this is a planning pause.

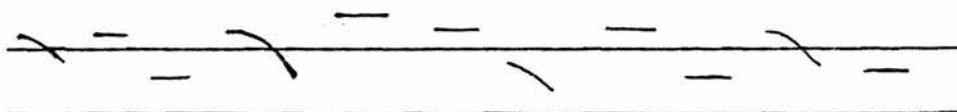
In most of the previous examples we have been dealing with samples of speech which look very like single tone units (apart from having rather more pitch movement than might be predicted by Halliday and Crystal cf. section 3). In other words the pause has occurred apparently in the middle of a single tone unit or at a

tone unit boundary (e.g. examples 7 and 8). When the pause has interrupted a tone unit in the previous examples, the tone unit has continued after a pause and then been concluded by a final pause at the tone unit boundary (cf. examples 3 - 8). But note the following example which typifies a pattern which is very common in spontaneous speech:

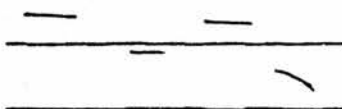
9. of course houses then were eh + room and kitchen



and single ends that was eh + more or less the +



trend of houses +



(22/13/1-2)

The pauses occur in the above example in the following places:

i) between a verb and its direct object i.e. were
eh + room and kitchen,

ii) between a verb and the adverbial phrase related to it i.e. was eh + more or less,

iii) between an article and its head noun i.e. the
+ trend of houses, and finally, at the end of the speech sample following houses +.

Throughout this example (and indeed the previous examples) there has been a lot more pitch movement than might be predicted, but if we look in particular between the lexical items single ends and that was, there seems to be a great deal of pitch movement on ends falling from fairly high to low, and this pitch movement fall is followed by a large step up to the high pitch level on that. This would seem to indicate a tone unit boundary, but there is no pause co-occurring with the tone unit boundary, instead, the pause occurs after the tone unit has started i.e. in the middle of the tone unit. We have already seen examples of this (cf. 3-7).

The first pause in the above sample occurs between were and room and kitchen. This pause occurs in the middle of a tone group (like 3-7 above) but it is not followed by another pause at the end of the tone group (which is what happens in the other examples) instead, the first tone group is completed and the second tone group is started before the speaker pauses again. Are we to regard these pauses simply as hesitation pauses or is there something else going on? It seems to me that the above strategy is a discourse strategy. The speaker seems to use this strategy as a means of retaining his 'turn'. If the speaker pauses at tone unit boundaries, and tone units are assumed to be correlated with 'completed' information at some level, then it would be

very easy for the Hearer to interrupt, forcing the Speaker to lose his 'turn'. On the other hand, if the Speaker pauses in the middle of a constituent such as prepositional phrase, or noun phrase or between a verb and its direct object etc. then the Hearer must wait for the unit to be completed. The strategy employed above could be illustrated as follows in general terms:

Beginning of Tone Unit 1 - PAUSE - Completion of TU 1,
Beginning of TU 2 - PAUSE etc.

^^ Such that the pauses always occur in the middle of a tone unit making it very difficult for the Hearer to interrupt.

This seems to be a very typical strategy in the data, and some examples are given to illustrate the point.

10. I suppose that's the same in most eh + areas

but although...

(22/9/4)

11. I would say Cannongate South Side I mean I

— — — — —

would term the Canongate + er South Side its

— — — — —

more or less the South area + of the city

— — — — —

(22/11/3)

12. when I went back + from the University like

— — — — —

from the Gymnasium roof looking + over into

— — — — —

Arthur Street where + you had three open...

— — — — —

(22/12/1-2)

All of the above examples are fairly lengthy and are taken from the transcript within a very short space of time, and all examples are from the same speaker. In order to show that this speaker is not the only speak-

er who adopts such a strategy examples will now be presented from other speakers.

13. But he was + young and eh + I can remember

the + the different kinds of bands too when I

was young like Harry Roy...

(20/10/4)

14. there was about twenty of them behind the bar +

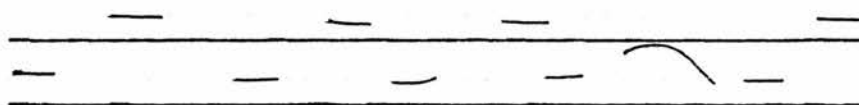
would rush up and sort of shovel + so many

glasses underneath the hot water tap and + start

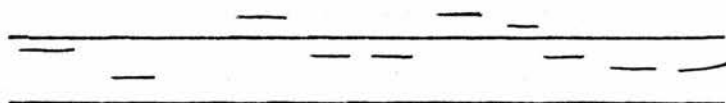
+ doing the rigmarole for + for Irish coffee +

(23/5/1-2)

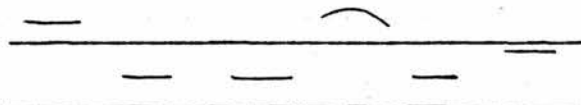
15. but that's the that eh + open ground across



from the + Deaconess hospital has eh



buildings there too you see +



(10/9/5)

Thus the strategy described above seems to be widespread amongst various speakers, and to be used regularly throughout the data when the speaker wished to maintain his turn. This strategy is not found in the questionnaire situation where turn-taking is conventional in terms of question / answer sequences, and it is rarely found in discussions of the photograph, since again the interviewee is being asked to carry out a specific task. The most common situation in which this strategy is found, therefore, is in spontaneous speech especially when there is a possibility of competition for the role of speaker.

The strategy which has been discussed throughout this section where the speaker chooses to pause in the middle of a tone unit in order to maintain his turn, means that the various functions of the pause may not

only be confused with each other, but they also seem to contradict each other, i.e. the speaker-maintaining pause will be very difficult to distinguish from the planning pause, and it seems to completely contradict the boundary-defining pause which Crystal proposes. We cannot hope to resolve these dilemmas completely but let us attempt to define the various functions of pause mentioned so far and see how far the different functions can be related to different realisations of pause.

A) There are two main types of hesitation pauses referred to above as:

- i) planning pause (realised by silence)
- ii) filled pause (realised by vocalisations such as em, er, eh, etc. which may or may not be accompanied by perceptible silence on either side)

These pauses can often be recognised easily as hesitation pauses since they occur between constituents such that the constituents are deemed to have been 'interrupted'. The planning pause functions as a breathing space in the stream of speech which allows the speaker to plan or replan the utterance which is still to come. The filled pause can be regarded as a special instance of planning pause. Note Livant (1963) who states that "filled pauses ... increase the speakers control of conversation, but decrease the quality of his production." Goldman-Eisler (1961) also notes that there

is a different output resulting from the habitual occurrence of these two phenomena but both investigators state that both of these types of 'hesitation pauses' function as 'breathers' to allow the speaker to plan ahead.

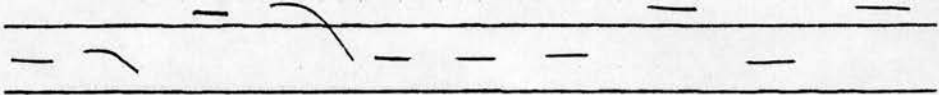
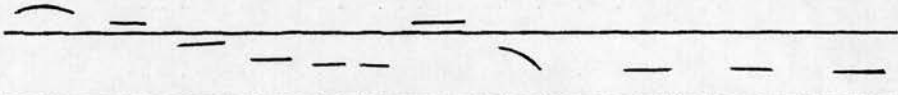
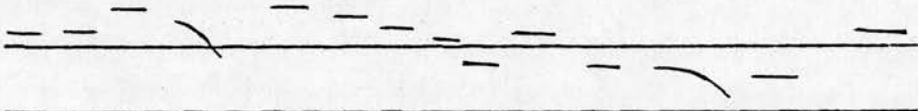
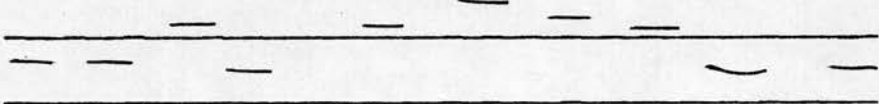
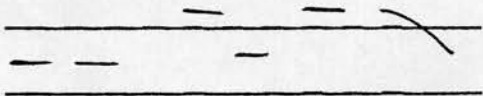
B) There are pauses of various lengths (see Crystal and chapters 5 and 7) which occur at tone unit boundaries. These pauses may be referred to as boundary-pauses, and their function is to delimit stretches of speech phonologically into units which we shall call tone units or tone groups. Boundary-pauses will occur between completed syntactic units of various sizes, but will not 'interrupt' syntactic or semantic units.

C) There is a third type of pause which functions in the discourse and we will refer to this type of pause as the speaker-maintaining pause. This pause occurs in the middle of a tone unit, normally either just before the final lexical item of the first tone unit (given a series of two) or just after the first lexical item of the second tone unit. Tone units, in the context of speaker-maintaining pauses, would therefore have to be defined according to internal criteria such as 'pitch movement cohesion' etc. (see chapter 6).

The function of the speaker-maintaining pause is for the speaker to retain his role of speaker by deliberately not pausing at syntactic or semantic boundaries in case he is interrupted and loses his role of speaker.

We therefore have three different types of pause which are operating on three different 'levels', or as part of three different systems.

- A) can be regarded as an 'accidental' pause which is not systematic in any way. This pause can occur at any time but will normally be recognisable as occurring in the stream of speech such that a syntactic or semantic unit is 'interrupted'. After the pause the tone group will then be completed and rounded off by a boundary pause.
- B) operates at the phonological level and will be treated as part of an independent system, such that pauses of different lengths will be seen to have different phonological functions. This pause occurs at tone group boundaries and tends to occur between whole syntactic and / or semantic units of some kind.
- C) operates at the level of speaker-interaction i. e. the speaker maintaining pause is a pause which operates in the context of a conversation between at least two participants. The typical features of this type of pause are a) that it occurs in the middle of a tone group, b) there will normally be several pauses following each other consecutively in the stream of speech, and c) the tone group following the first pause is not marked by a boundary pause, i.e. the pause follow-

- but eh + you know it + it + although they don't
- ii) 
- seem to have a very good eh + what's the word
- iii) 
- reputation + in America singles bars from what
- iv) 
- I could gather + this this one was nice + it
- v) 
- was eh + very pleasant +
- vi)  (23/5/5)

The above sample of spontaneous speech is a typical example of spontaneous conversation illustrating some of the problems that can arise when we try to analyse spontaneous speech. This stretch of speech contains many examples of planning phenomena which make an analysis of intonation into tone groups and tonics very difficult. Below is a list of the various examples of the different kinds of planning phenomena which are in evidence in this particular sample of speech:

- a) repetition: line (ii) you know it + it +
line (v) this this one was nice +
- b) false start: line (ii) you know it + it + al-
though they don't ...
- c) lexical filler: line (ii) + you know it +
line (iii) + what's the word
- d) filled pauses: line (i) + eh +
line (ii) + but eh +
line (iii) a very good eh +
lines (v-vi) it was eh +

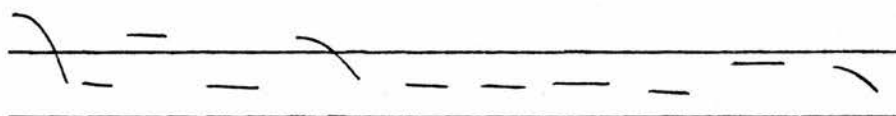
The notion of filled pause has been extended here to include short phrases such as a very good which are not completed but bounded by the lexical filler eh. It is obvious that there is in fact a continuum between filled pauses and lexical fillers. If we adopt the standpoint that a filled pause merely contains a meaningless vocalisation then we would have to include non linguistic vocalisation in this category such as sneezing, belching, giggling etc. Instead we will identify incomplete phrases of 2-3 words containing the filler eh or em as filled pauses, and other phrases containing items without the filler eh will be referred to as lexical fillers.

The above examples are the obvious ones. We could take a step further and suggest that perhaps lines (v-vi) it was eh + very pleasant is in fact a semantic

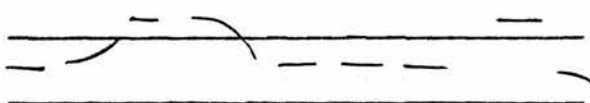
repetition or reformulation of the previous phrase in line (v) this this one was nice. Such strategies of repetition of phrases or lexical reformulation being perhaps extended examples of the phonetic correlates of the strategy which we are referring to as 'planning'. All these means exemplified above allow the speaker to pay less attention to scanning his present output and more attention to formulating his next output.

Further examples of similar phenomena are as follows:

17. oh apart from once when we went we found em +

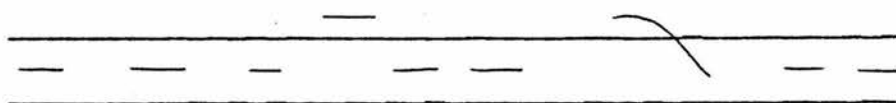


an Irish bar in San Francisco

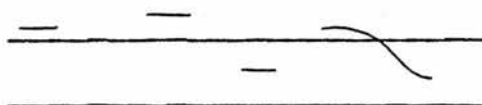


(23/4/4)

18. well there's a list in your + drawer + in the



ches- chest of drawers +



(23/9/4)

19. well + my cousin's wife was + doing her best

— — — — —

to show me + you know she's she showed me the

— — — — —

Mormon cathedral

— — — — —

(23/12/4)

All of the above examples (17-19) contain planning phenomena of the various kinds mentioned in connection with example 16. The types of phenomena exemplified are enumerated below:

- | | |
|--------------------|---|
| a) repetition: | 18. in your + drawer + in the
ches- chest of drawers |
| | 19. to show me + you know she's
she showed me... |
| b) false start: | 17. when we met we found |
| | 18. in your + drawer + in
the ches-... |
| | 19. to show me + you know she's
she showed me ... |
| c) lexical filler: | 19. you know |
| d) filled pauses: | 17: em + |

- e) planning pauses: 18. in your + drawer + in the
 19. my cousin's wife was + do-
 ing her best
 19. to show me + you know

Thus from the above exemplification we can see that there are many different realisations of planning phenomena in the data. Let us look at some of the above examples in detail in order to see how planning phenomena affects the identification of tone groups and tonics.

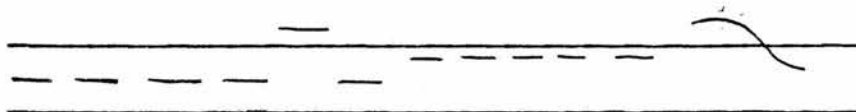
We have already mentioned the problem caused by lexical fillers in section 4.3. The lexical filler could be regarded as a special case of filled pause perhaps, but instead of being a vowel or some other meaningless vocalisation, we have lexical items used as 'place-holders' allowing the speaker to maintain his role while he plans ahead or searches for a lexical item.

In the examples above we have a lexical filler surrounded by pauses and containing some pitch movement (cf. 16 line (ii) + you know it +). This example might suggest that lexical fillers should be treated as separate tone units. But in example 19 the lexical filler is preceded by a pause but not followed by one, instead, it seems to be part of the tone unit following it, so much so that the lexical filler is the element containing the maximum pitch movement in the unit (you

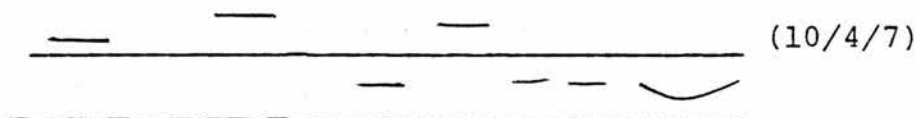
know she's she showed me the Mormon cathedral). Is the lexical filler to be joined to a tone unit, to be regarded as a separate element, a tone unit in its own right, or to have no status as a tone unit at all? Perhaps we should allow any of these realisations but assign it a very specific function. We might say therefore that the lexical filler has a unique function in the discourse as a realisation of planning phenomena which allows the speaker to maintain his role of speaker while searching for a lexical item or planning ahead in the discourse. The lexical filler, then, may occur anywhere in the discourse (since one cannot predict where one might 'lose' a word) i.e. either forming part of a tone group, or in between tone groups seeming to belong to either, or as a tone group in its own right. Other items such as see, you see, well, yes can also be used in this way.

False starts and repetition cause similar problems in the analysis of intonation, therefore we will examine these phenomena together. It is often the case that false starts involving repetition cause difficulties in identifying tone units, note the following example:

20. and it was a garden separated the Brown's +



Brown's Street from Salisbury Street



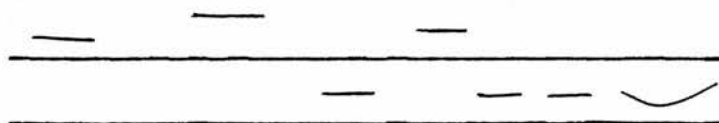
In the above example, pitch movement criteria would lead us to place the tonic on the lexeme street. But if we look at the stretch 'Brown's Street from Salisbury Street', or listen to it isolated from the rest of this stretch of speech, street might be interpreted as somehow contrative (i.e. with the meaning not Brown's 'square' or 'place', but Brown's 'street'. This stretch is one which has been replanned. The speaker probably intended to say 'and it was a garden separated Brown's Street from Salisbury Street', but something went wrong and the speaker produced 'the Brown's'. He therefore stopped, paused and replanned the same stretch and produced it again, repeating the element Brown's. Because the word Brown's was produced initially with a great deal of pitch movement (presumably to contrast with Salisbury), when the same element is produced for the second time it receives very little pitch movement and the pitch movement seems to have been displaced on- to the element Street.

There is a lot of discussion in the literature⁽¹⁾ where it is suggested that when a constituent is repeated the tonic moves one place to the left. Repetition involved in false starts then, seems to support the

first part of this notion (i.e. that the tonic moves) but since the original tonic was on Brown's (presumably of Brown's street) and the repetition causes the tonic to move to the element street, we must extend the directional notion of the movement such that we allow for the tonic to move in any direction away from the original element. In the case of example 20 the tonic moves one place to the right. In the literature the tonic moves to the left presumably because the repeated element is the rightmost element. Therefore we might make a generalisation as follows:

In a replanning situation involving repetition due to false starts, if the element which is repeated contains a lot of pitch movement when produced initially, this pitch movement will tend to move away from the repeated element onto another element in the replanned stretch of speech. The element receiving the pitch movement may be to the right of the original element as in --

20. Brown's street from Salisbury street

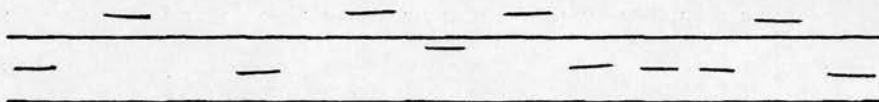


where the original pitch movement was on Brown's and it has moved to Street.

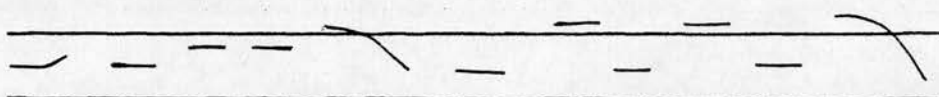
Another example of the pitch movement being displac-

ed to the right of the original movement can be found in the following example:

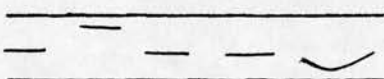
21. But this + it's alright looking on a picture



but it's pretty sad + it's pretty sad to live +



in places like that

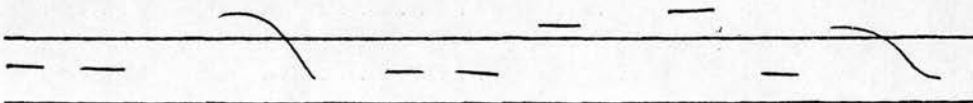


(9/3/6)

In example 21, the initial maximum movement is on the element sad, but this movement is displaced onto the element live.

Alternatively, the element receiving the eventual pitch movement may be to the left of the original as in example 18 where maximum pitch height is moved onto the element chest although drawers still retains some movement.

18. in your + drawer + in the ches- chest of drawers +



(23/9/4)

We can see from the above examples, then, that planning phenomena tends to create anomalous tonics. We must take such phenomena into consideration, therefore, when attempting to define tonics and tone units in spontaneous speech. Not only do we get anomalous tonics but we also get 'tails' and 'heads' being separated from their tonics by planning pauses such that we get anomalous tone units being created because of planning phenomena as well. In the above example (21) for instance, the phrase in places like that is separated from the rest of the unit by a pause, yet there is hardly any pitch movement at all on any of the elements in the phrase. Are we then justified in identifying a separate tone unit?

When dealing with spontaneous speech it is essential to be constantly aware of planning phenomena since it occurs frequently in the data. Such phenomena makes the analysis of spontaneous speech hazardous, to say the least, since anomalous tonics and anomalous tone groups are constantly being created. I think planning phenomena should be treated simply as a realisation of performance variables. The phonetic correlates of planning phenomena should not play an active part in the definition and delimitation of tonics and / or tone units.

4.6 Pitch movement in the data

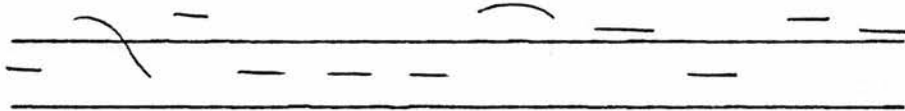
The above examples used to discuss pause phenomena

already indicate some of the problems which have to be recognised in the area of pitch movement. It is obviously the case if we look at some of the above examples (especially sentences 4-9) that even when there is an interruption in the stream of speech due to a hesitation or planning pause, the stretch of speech delimited by planning pauses contains more than one major pitch movement. This 'movement' is not always in the form of continuous movement, if we look at the examples above we find that in the majority of the data, there tends to be occurrences of jumps in pitch rather than continuously moving pitch on the whole. There seems to be an overall pattern where there is a baseline established by the unstressed syllables from which 'jumps' in pitch then depart. The only noticeable exceptions occur when we have monosyllabic lexical items occurring at the end of a tone group especially when the monosyllable is composed mostly of voiced segments, e.g. if the word days occurs finally in a tone group we tend to perceive pitch movement whereas if the word distant occurs finally in the tone group we tend to perceive a 'jump' in pitch. This is mainly due to the fact that days is voiced throughout. Therefore if the final lexical item is monosyllabic and contains a peak, we will have continuous movement back to the baseline. If we have a polysyllabic word with a voiceless segment in the middle separating the syllables we will tend to get

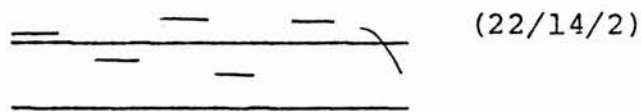
a pitch jump from the peak to the baseline.

Most of the examples above have involved pauses, which have confused the identity of the tone group. Let us now examine some stretches of speech which are delimited by boundary pauses and which do not contain planning or speaker-maintaining pauses. The following are some examples:

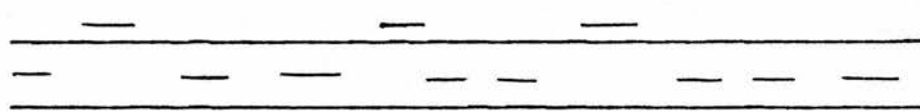
22. in fact I would say in these days that if you



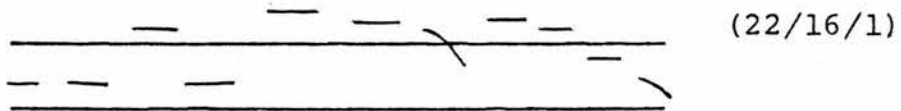
had an inside toilet



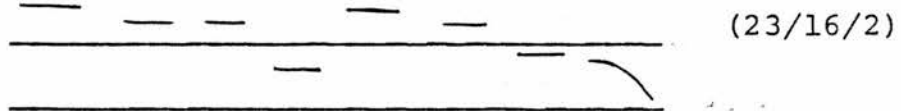
23. a thing that comes easy you don't get the same



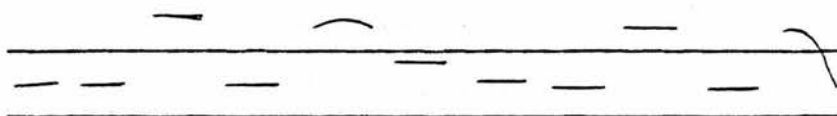
eh enjoyment out of it definitely



24. that came as the biggest surprise

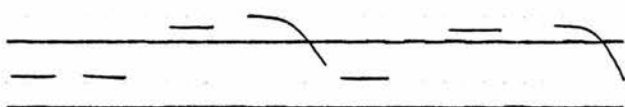


25. because I lost two dollars and that was it



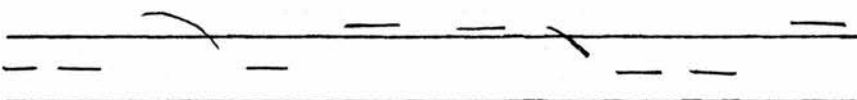
(23/13/5)

26. did the old man have three sons

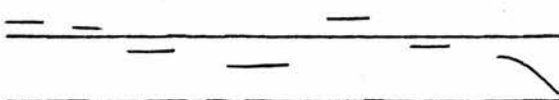


(AD/3/4)

27. I was quite impressed with it its the first

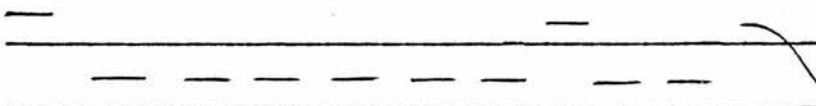


holiday's we've had up there



35/6/6)

28. I thought I was getting insulted you know



(31/3/1)

The above examples are all surrounded by boundary pauses and are not interrupted by planning pauses (although there is one example of a filled pause in example 23). All of the above examples, then, would presumably be analysed by Crystal as tone units, but all of the examples contain more than one maximum deviation from the baseline. How is one to decide, therefore, which

'peak' is to be regarded as the nuclear movement of the tone unit. Crystal's second criterion of 'change of direction' might be used to sub-divide the tone units as they stand at present, but then example 23 would be sub-divided as follows:

(a thing that comes) (easy you) (don't get the same
eh en-) (joyment) (out of it) (definitely)

where not all of these units can be said to have grammatical or semantic coherence. Thus even when tone units are not interrupted by planning pauses there still seems to be a conflict between the various cues proposed by Crystal for identifying tone unit boundaries.

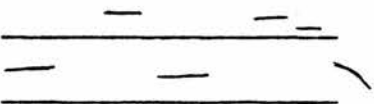
The above examples might be divided into units according to grammatical or semantic criteria, giving the following units for example 28:

28. (I thought) (I was getting insulted) (you know)

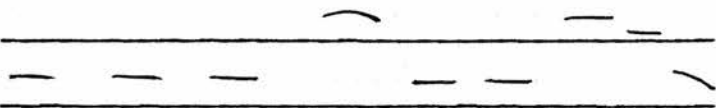
Thus example 28 contains three units, each containing only one peak of prominence according to the auditory transcription. If all of the above examples are divided up such that each tone unit contains only one peak of prominence, however, then the domain of the tone unit becomes very small indeed. See for example the following

samples of speech where all of the examples have been taken from the speech of the same speaker within a very short section of his spontaneous speech to demonstrate that such units are typical, and occur frequently in the data within a short space of time:

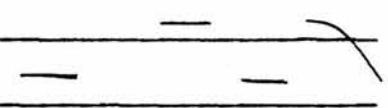
29. well in that area

 (22/12/3)

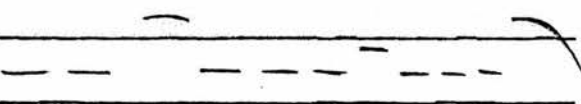
30. when there were hundreds of families

 (22/12/3)

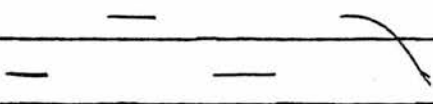
31. they're looking now

 (22/12/3)

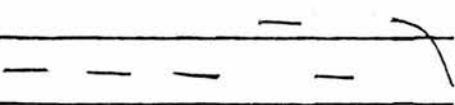
32. it was hardly any area at all

 (22/12/4)

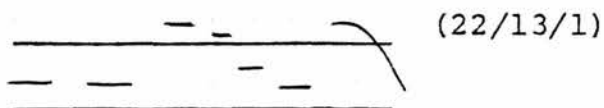
33. well when you're young

 (22/12/5)

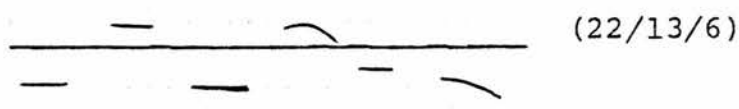
34. but when you see it now

 (22/12/5)

35. and the area itself



36. they put them in of course



In examples 29-36, the deviations from the baseline occur on almost every stressed syllable, and all of these deviations seem to occur at the same level on the stave i.e. level 4. Thus the deviations (or peaks of prominence) all seem to reach the same height according to the auditory transcriptions. Perhaps it is the case that despite the fact that the peaks all reach the same height, they do not all possess the same amount of pitch movement. Therefore perhaps the syllable with the most pitch movement is to be regarded as the 'nuclear' syllable or tonic of the tone unit?

If we examine the items which contain pitch movement rather than jumps in pitch we find that in all cases, the item containing pitch movement is the final stressed item in the tone unit. This item is stressed therefore it must deviate from the baseline, but since it seems to be the case that all speakers return to the baseline in a normal declarative narrative style, the speaker must use pitch movement in order to get back to

the baseline. There are no following syllables which can be uttered on the baseline to indicate that the speaker has returned to the baseline. If we examine examples 29, 30 and 36, there are unstressed syllables which follow the stressed syllable, and the movement back to the baseline is 'spread' over the unstressed syllables. There seems, therefore, to be a phonetic environment which determines whether the pitch moves continuously from level 4 downwards, or whether the pitch jumps from level 4 to level 2 or 1. It may also be the case that the movement at the end of a unit is perceived as being 'lower' and therefore more 'nuclear' than movement which occurs in the middle of a tone group.

In order to examine whether the final movement is perceptually more prominent or is merely phonetically predictable according to the environment, a series of experiments were set up to test which of these two hypotheses would be supported by independent judges.

Chapter 5.

5.0 Introduction

In chapter 3 we discussed two of the main theories which had been proposed for British English intonation i.e. the theories proposed by Crystal (1969), and Halliday (1963, 1970). Both of these investigators maintain that in a normal, unmarked tone unit or tone group there is one main peak of prominence. Halliday (1967) believes that an unmarked tone group will be co-terminous with one information unit, which in its unmarked form, will be co-terminous with a clause. Halliday maintains, therefore, that in an unmarked clause, there will be only one peak of prominence. Halliday and Crystal describe such a peak as follows:

"Within each information unit, one part is selected as prominent; this is the tonic... The function of the tonic is to form the focus of information..."
(Halliday, 1970, p. 40)

"Every tone unit contains one and only one nucleus, or peak of prominence, expounded by one of a finite number of contrasting pitch glides or sustentations on the accentual syllable of the most prominent word."
(Crystal, 1969, p. 209)

"... placement of the nuclear tone... is generally said to signal the point of major interest or most

information in the tone unit. Normally, the nucleus falls on the last lexical item in the tone unit." (Crystal, 1969, p. 263)

Other investigators in the field of intonation have also proposed this notion of one main peak of prominence per 'unit', cf. the following:

"... the place where the nuclear, stressed syllable of a primary contour may fall is in general determined not by the system of pitch contrasts but by the element of the lexical hierarchy." (Pike, 1965, p. 107-8)

"... a minimal unit, resulting from the division of speech by intonation and meaning is a syntagm, which the academician Scerba (1957) defined as 'the phonetic whole expressing one unit of meaning' ... Any syntagm in Russian has certain dynamics of fundamental frequency change, as well as changes in intensity and duration. Besides its dynamics, a syntagm has, as a rule, one maximal value of each parameter... A syntagm, as a unit of paradigmatic opposition, can be either the whole syntagm or its called 'information centre' i.e. a part of a syntagm located on the most important word of a sentence which is usually found at the end of syntagm." (Svetozarova, 1975, pp. 499-503)

All of the above quotations seem to correlate a maximum peak of prominence with the 'focus of information', this peak of prominence being realised phonologically as the maximal value of the phonetic parameters, fundamental frequency, intensity, and duration; cf. Halliday who says:

"Each point of information focus is realised as a tonic component;... The initial syllable in the tonic component, the 'tonic syllable', is phonologically prominent, this prominence being primarily a matter of pitch (pitch movement, not pitch height) and secondarily one of duration and intensity." (1967, p. 203)

The examples presented in section 4.6 of chapter 4 raise problems for the above theories if all peaks of prominence in each utterance which are marked auditorily on the transcription as being equal, are perceived as equal. Examples 29-36 contain two peaks auditorily, but if Halliday is correct in assuming that pitch movement determines prominence rather than pitch height, then in the aforementioned examples, the final peak would be heard as prominent in most of these examples since the final peak is realised as a continuously moving fall in pitch, whereas the preceding peaks are generally realised as jumps in pitch. It was suggested in chapter 4

that the reason for such variation may well be the phonetic content of the utterance i.e. when the final stressed syllable is a monosyllable, there will be a continuously moving fall; when the final stressed syllable is followed by other unstressed syllables then the pitch movement will be distributed over the syllables in such a way that the realisation may be a jump in pitch.

Peaks which occur in the middle of an utterance can be realised as a jump in pitch because there are syllables following the peak which may be uttered on the baseline. An utterance-final stressed peak occurring on a monosyllable has no other way of returning to baseline except by falling continuously.

There are therefore two conflicting hypotheses involved here;

- 1) Halliday's hypothesis that pitch movement will determine prominence
- 2) the hypothesis which would be proposed in this thesis that peaks which are transcribed as equal, are perceived as equal since the difference in their realisation is attributable to phonetic context.

If examples 29-36 are analysed in the light of the first hypothesis, then each example would be seen to contain one peak which is more prominent than another. If the same examples are analysed in the light of the second hypothesis, then they would be seen to contain

two equal peaks of prominence. If the utterances are seen to have two equal peaks of prominence, then examples 29-36 pose problems for Halliday's proposal that the tone group in its unmarked form is co-terminous with the clause. The examples would have to be divided into two units, each containing one peak of prominence. The tone group would then become a much smaller unit than the unmarked 'clause' which Halliday proposes, taking the shape of a unit of meaning approximating more closely to the syntactic unit 'constituent'.

Crystal and Halliday above equate the tone group with the unit of information but there are many problems involved in dividing up an utterance into 'units of information' since even the definite article the has the meaning 'definite' as opposed to 'indefinite'. For this reason investigators normally adopt the position that the 'tonic', 'nuclear tone', 'nuclear stressed syllable' or 'intonation centre' marks the domain of the 'tone group', 'tone unit', 'primary contour', and 'syntagm' respectively. There is a plethora of terms in the literature all of which are used to refer to similar elements. I will refer to the single, maximum peak of prominence within a unit as the 'tonic' or 'tonic element', and the unit of its domain as the 'tone unit', since each tone unit is said to contain only one tonic.

In the literature, prominence or stress is said to be realised by a combination of phonetic parameters (see

chapters 1 and 2 for a more detailed discussion). These parameters are commonly agreed to be: fundamental frequency (F_0), intensity and duration. The maximum peak of prominence within an utterance should therefore be signalled phonetically by maximum F_0 measurement, maximum intensity and maximum duration. In chapters 1 and 2 the problems involved in including duration as a relevant parameter in the study of intonation were discussed briefly (see especially 2.1). The majority of experiments which have examined duration as a cue for stress identification have involved the opposition of one stressed syllable with one unstressed syllable in bisyllabic words such as digest and object which can be realised with main stress on either syllable. If we are to arrive at a 'maximum' measurement for the duration parameter in order to identify the 'maximum peak of prominence', is the longest word in the utterance to be regarded as the maximum (i.e. the word with the maximum number of syllables), or the longest monosyllable, or the longest vowel?

If the longest monosyllable is said to be the element with maximum duration, do we mean the longest monosyllable compared with other monosyllables within the utterance, or do we mean that a particular monosyllable is longer than 'usual', and may therefore be said to have the maximum duration of its allophones? It should be noted that 'length' plays an important part

in distinguishing one phoneme from another. For example Sharf (1964) states that the phonemic opposition 'voiced' versus 'voiceless' is often realised as an opposition of length such that a voiced fricative is often distinguished from a voiceless fricative by length alone and not by the presence or absence of voicing.

The problems involved when including a duration parameter as a cue for prominence identification at the level of intonation seem complex and numerous. Duration was therefore not included as a relevant parameter for identifying maximum prominence within an utterance.

A series of pilot experiments was set up in order to test whether judges who were familiar with the notion 'tonic', would identify tonics according to the aforementioned criteria of phonetic maxima.

5.1 Experiment 1

Hypothesis: the hypothesis being tested in this experiment is that the cumulation of phonetic maxima when present in an utterance will be identified by judges as the realisation of the tonic element. When the phonetic maxima are not cumulated on one element, but spread over several elements in the utterance, the judges will be forced to choose between the various phonetic maxima.

5.1.1 Method

The first experiment was constructed to examine

three phonetic parameters and their contribution to the identification of maximum prominence. These three parameters were (A) maximum pitch height; (B) maximum pitch movement; and (C) maximum intensity. The F_0 measurement was divided into two separate parameters because the maximum pitch movement did not always co-incide with the maximum pitch height. Since Halliday's hypothesis outlined in 5.0 maintains that movement is more prominent than height, the two parameters were measured separately. The parameters were measured in order to establish a maximum value for each parameter within a sentence. The syllable with the greatest height in cycles per second within a sentence was defined as the syllable with the maximum pitch height for that sentence. When there are two or more syllables within the sentence all reaching the same height, and when that height is the greatest height attained within the sentence there are two or more maximum pitch heights. When there are two equivalent amounts of movement within the sentence, and that amount of movement is the greatest amount of movement within the sentence, then there are two maximum pitch movement measurements, similarly with intensity. E.g. given sentence 14 in Appendix 1 How many children have you though, how, children, and have, all have the same amount of movement, and there is no greater amount of movement within the sentence therefore all three items can be said to have maximal pitch movement. How

is the highest item in the sentence, and children has the greatest amount of intensity within the sentence, therefore the sentence is defined as follows for the purposes of the experiment:

HOW(AB) MANY CHILDREN(BC) HAVE(B) YOU THOUGH

All three parameters were combined on the output of a four track Mingograph. There are occasions when it is difficult to tell which syllable of a word is the syllable which has the maximum, but since the subjects were asked to identify the tonic word of a sentence, the problem was not relevant.

It seemed reasonable to suppose that the phonetic cues of one dialectal area may not always be interpreted in the same way by speakers of other areas, since the physical properties of intonation contours differ from one area to another(cf. descriptions of the form of intonation patterns in various languages -- cf. Adams, 1969; Bowen, 1956; Abe, 1957/8; Delattre, 1965; t'Hart, 1973; Cohen, 1967; Svetozarova, 1975 and Esser, 1978). In order to test this hypothesis, a total of 29 judges were selected from different accent areas. Phonetic experience was also taken into consideration to test whether this had any effect on the judgments. Appendix 4 was therefore used as the front page for all of the experiments to elicit the inform-

ation needed from the judges taking part.

In order to obtain the data for the experiments a group of native Scots speakers with no formal knowledge of the area of intonation was asked to read a set of 15 sentences out of context. From these readings, 20 sentences were selected on the basis of the combinations of the phonetic maxima (PM) occurring in the sentences. The three phonetic maxima did not always cumulate on the same syllable but were often divided and combined in different ways. The sentences were selected from sentences read out of context so that all the lexical items were new and not given in the expectation that the PM might then be divided between lexical items as indeed happened.

The judges were asked to select 'the tonic word for each sentence', i.e. 'the most prominent word in the sentence'. If more than one word was heard to be 'most prominent', then more than one tonic could be selected. Given sentence 3 in Appendix Have you change of a pound, with the maxima distributed as follows:

HAVE YOU CHANGE (AC) OF A POUND (B)

a judge who selected only one 'tonic' would presumably be forced to 'choose' either change or pound, and according to the choice he made, his preference for

a particular phonetic criterion would be made clear. Judges could also decide that each element was equally important, although signalled in a different way, and therefore select two 'tonics', change, and pound.

Out of 100 sentences read by these Scots speakers, only 24 sentences contained a cumulation of all the phonetic maxima; 32 contained (B) on one syllable and (AC) on another; 17 contained each maximum on different syllables having the combination (A) (B) (C); 15 contained the combination (C) (AB) and only 9 contained the combination (A) (BC). The remainder combined the maxima in various ways. These measurements indicate that there is a cumulation of the phonetic maxima being measured for this experiment on only approximately 25% of the sentences measured. The rest of the sentences contain a spread of maxima realised in different ways. The above measurements indicate that it is much more common to have movement on one syllable and height co-occurring with intensity on another. In other words it seems to be true for Scots speakers, height co-occurs more readily with intensity than movement does.

From the above possible combinations the following were selected to make up the 20 sentences:

(ABC)	7 times	(cumulation of all 3 PM)
(A) (B) (C)	4 times	(competition between each PM)

(B) (AC)	2 times	(one PM competing with remaining two)
(C) (AB)	2 times	(one PM competing with remaining two)
(A) (BC)	once	(one PM competing with remaining two)
(ABC) (AC)	once	(one cumulation plus a combination)
(AB) (BC) (B)	once	(one spread PM in various combinations)
(AB) (AC)	once	(one spread PM in combination)
(AC) (BC)	once	(one spread PM in combination)

The results are presented in Appendices 1 and 2 , retaining the order of presentation of the sentences. The results are also presented in Appendix 3 , with sentences containing similar lexical items placed together to allow easy comparison between sentences with similar lexical content containing a different spread of phonetic maxima.

The number of judgments appears to vary from sentence to sentence because some judges decided to select one tonic while others decided to select more than one tonic. It was not the case that judge X consistently selected one tonic while judge Y consistently selected two tonics. The judgments varied for each judge from one sentence to another, thus the number of judgments

also appears to vary from one sentence to another. For example, given the following sentence:

THE OLD(A) MAN(C) ASKED IF SHE HAD POSTED(B)
THE LETTER

Judge X might select letter as the only tonic; judge Y might select asked, and letter as two equally prominent tonics; and judge Z might select old, asked, posted and letter as four equally prominent tonics. The results would then look like the following (with the number of judges selecting more than one tonic given in brackets on the lower line):

THE	OLD(A)	MAN(C)	ASKED	IF	SHE	HAD
	(1)		(2)			
POSTED(B)	THE	LETTER				
		1				
(1)		(2)				

For a sentence such as Have you change of a pound, where the same three judges X, Y and Z all select pound as the one and only tonic, the results would look like this:

HAVE YOU CHANGE(AC) OF A POUND(B)

There are no bracketted readings here because all the judges chose one tonic only, realised on pound.

5.1.2 Sentences containing the cumulation (ABC)

If we first of all examine sentences 2, and 9, where each sentence contains one syllable with the cumulation (ABC), then the results are seen to support the hypothesis that the tonic is realised by a cumulation of maximal measurements of the three phonetic parameters under consideration. The majority of judges select the element with the cumulated maxima as the tonic.

2. THREE (AC) SONS (ABC)

2 24

(3) (3)

9. THREE (ABC) SONS

27 2

Similarly, sentence 16, has the cumulation of maxima on one syllable, and again the majority of judges select that element as the tonic, cf.

16. DO YOU WANT TO PLAY (ABC) WITH ME

(1) 27 (2) (1)

According to the results so far the judges seem to

to be selecting the cumulation (ABC) when it is present as predicted. If we now examine sentences 19 and 5 we can compare two sentences which have the same lexical content, but different phonetic cues.

19.	HOW	PLEASED (ABC)	THEY	WERE	WHEN	THEY	HEARD	IT	
		27						1	
		(1)						(1)	
								(AB)	
5.	HOW	PLEASED (AC)	THEY	WERE	WHEN	THEY	HEARD	IT	
		9					10		
		(9)					(1)	(10)	

Again the results are exactly as predicted with the cumulation (ABC) being selected by 27 out of 29 judges in sentence 19, and in sentence 5 the judges are split almost evenly between the two competitive maxima, (AC) on pleased, and (AB) on heard. 9 judges select pleased, 10 judges select heard, and 10 judges select both words as equally prominent.

So far then the cumulation (ABC) is selected by the majority of judges as the realisation of the tonic element. This is not true for all sentences containing the cumulation (ABC). Sentences 4, 13, and 20 contain the cumulation (ABC) and no other maximal measurements, yet the cumulation (ABC) is not selected by the majority of the judges as the realisation of the tonic element. Let us compare the sentences which did have the cumulation (ABC) selected as the tonic by the majority of judges, and the sentences which did not.

(ABC) selected by majority as tonic:

2. THREE (AC) SONS (ABC)
 2 24
 (3) (3)

9. THREE (ABC) SONS
 27 2

16. DO YOU WANT TO PLAY (ABC) WITH ME
 27
 (1) (2) (1)

19. HOW PLEASED (ABC) THEY WERE WHEN THEY HEARD
 27
 (1) (1)

IT
 1

(ABC) not selected by majority as tonic:

4. THERE (ABC) IS MY HOUSE
 12 14
 (3) (3)

13. HE PLAYED (ABC) SQUASH ON MONDAYS WEDNESDAYS
 9 7 6
 (1) (5) (1)

AND FRIDAYS
 3
 (3)

20. HE LOST (ABC) EACH OF THE GAMES HE PLAYED
 4 1 17
 (6) (1) (6) (2)

It might be argued that all of the sentences containing the cumulation (ABC) not selected as tonics by the majority, have the cumulation in sentence initial position, but then so do 2 out of the 4 sentences whose cumulation (ABC) is selected as tonic, therefore there can be no difference between perceived prominence in

sentence initial position and in sentence final position. Why then are other elements in sentences 4, 13, and 20 perceived as prominent, when they are not realised by the cumulation of phonetic maxima?

In sentence 4, the last lexical item house is selected as the single tonic by 14 judges out of 29. This result supports the position held by Crystal, Halliday and others that the rightmost lexical item is the unmarked tonic. In sentence 4, house is the rightmost lexical item and despite the fact that it is not realised by any phonetic maximum, it is selected as the single tonic element by 14 out of 29 judges. If the phonetic realisation of the two elements selected as tonic in sentence 4 is examined, we find that house falls from 300-225cps, whereas there rises from 250-375cps. There not only contains the maximum movement (almost twice as much movement as realised on house) but there also reaches the highest pitch value (and contains the maximum intensity).

Throughout these experiments fundamental frequency is being treated as equivalent to perceived pitch. In chapters 1 and 2, several other factors were noted as being relevant to pitch perception, notably intensity. In this case the intensity measurement supports the fundamental frequency measurement. In chapter 2 fundamental frequency was shown to be by far the strongest cue for pitch identification, and the majority of investigat-

ors examining the acoustic correlates of intonation have concentrated on the fundamental frequency measurements. Despite all of this, it is not yet known whether pitch movement as it is realised in speech, is perceived more readily or is perceived as more prominent when the pitch is falling than when it is rising. It is not known whether pitch is perceived to be more prominent when it is low in the pitch range or when it is high in the pitch range. Pitch might be perceived as more prominent when it occurs on the rightmost lexical item than when it occurs on an earlier item (see Hadding-Koch and Studdert-Kennedy, 1964).

All of the above questions are as yet unanswered with respect to pitch phenomena co-occurring with linguistic phenomena. Throughout these experiments, therefore, pitch is equated with fundamental frequency measurement as suggested in chapter 2, and when problems such as the above occur, they will be pointed out and possible solutions suggested. Thus in the case of sentence 4, we might suggest that the pitch movement on house is perceived as equally prominent as the pitch movement on there because house is the rightmost lexical item. It might be argued of course that the tonic element is not being identified according to the realisation of phonetic parameters at all.

Sentence 13 looks very complex. Of the judges who selected a single tonic, 9 selected played; 7 selected

squash; 6 selected the constituent Mondays, Wednesdays and Fridays; and 3 selected Fridays. Of the judges who selected more than one tonic, 4 judges selected both squash and either Fridays or the whole constituent Mon-days, Wednesdays and Fridays; 1 judge selected both squash and played as the two tonics. If we examine the acoustic results for this sentence in Fig 1 played obviously contains the greatest intensity by far, is the highest word in the sentence and has more movement than any other item in the sentence. The intensity and F_0 print-outs are as follows:

Fig. 1



If we divide sentence 13 into two parts -- he played squash and on Mondays, Wednesdays, and Fridays, on the basis that 4 judges selected one element out of each part

as equally prominent, then the results may be accounted for in terms of the rightmost lexical item of each unit. In the first unit, 7 out of the 16 judges who selected some part of this unit as the most prominent, selected squash as the tonic. Squash could be regarded as the rightmost lexical item of this unit. The remaining 9 judges selected the cumulation (ABC) as the tonic. Out of the 9 judges who selected some part of the final unit, 6 selected the unit as a whole being unwilling to choose one part of the unit as more prominent than another part. The remaining three selected the rightmost lexical item of the unit (and of the sentence as a whole) as the tonic element. Thus the apparently complex results of sentence 13 can be accounted for if we accept the fact that the rightmost lexical item is perceived as prominent. In the case of sentence 13, the prominence is not obviously related to the phonetic parameters, therefore it seems to be the case that the phonetic parameters alone do not signal the tonic element.

In sentence 20, again the results seem complex, but since 6 judges select both lost and games as being equally prominent, and since there are two clauses in this sentence, we might be justified in dividing the sentence into two units as in sentences 4, and 13. Only 4 of the judges who select a single tonic select the cumulation of phonetic maxima as the tonic, 17 of

the remaining 18 select games which is not the rightmost lexical item of the sentence. No judge selects played, the rightmost lexical item, as the tonic element of sentence 20. Perhaps since lost and games have already been mentioned in the utterance, the fact that the games have been played is assumed as 'given' in Halliday's sense (i.e. recoverable from the preceding discourse (1967)). If this is the case, perhaps it is not unreasonable to suppose that the 'tonic' moves from the rightmost lexical item to the rightmost lexical item containing the focus of information. We seem to be moving away from the notion that the tonic is marked by the cumulation of phonetic maxima into the realms of discourse and information structure.

Let us examine the phonetic correlates for the two items lost and games which attract the majority of the votes between them. Games is a rise-fall measuring 250-285-200cps, and lost is also a rise-fall measuring 200-450-275cps. Thus lost certainly contains maximum realisation of pitch height, and in terms of fundamental frequency also contains the maximum pitch movement (maximum intensity is also realised on lost). However, games falls lower than lost, and is the longest stressed word in the sentence (games being 0.32 secs long; played being 0.24 secs long; and lost being 0.18 secs long), therefore this combination of phonetic features plus the fact that played may be treated as 'given' may well account

for the fact that the majority of judges selected games as the tonic of the sentence, rather than lost which is realised with the cumulation of maxima.

The problems involved in using duration as a parameter in intonation studies can be exemplified here. All of the words in sentence 20 are monosyllables, therefore the problem of comparing monosyllabic words with polysyllabic words does not arise in this instance, but how valid is a comparison between games and lost? Games consists entirely of voiced segments including a 'long' vowel (cf. Sharf, 1971 and Peterson & Lehiste, 1960) and voiced final consonants, whereas lost consists of a short vowel and two final voiceless segments. No pitch will be perceived on the voiceless segments and the short vowel will mean that the hearer has less time to perceive the pitch movement realised on it than when pitch movement is realised on a long vowel. Does this mean that all stressed lexical items which consist of voiced segments and long vowels are potential tonics when compared with lexical items consisting of short vowels and voiceless segments? Or should the item games, a token of which appears in sentence 20 be compared with other tokens of the same type in order to establish its 'normal' length, thus in turn establishing an 'extra long' length? There are many complex problems involved in measuring duration outwith the single lexical item, I therefore propose to leave this problem alone.

If we compare sentence 12 with sentence 20, the results appear to be anomalous:

20.	HE	LOST(ABC)	EACH	OF	THE	GAMES	HE	PLAYED
		4			1			17
		(6)			(1)			(6)
								(2)
12.	HE	LOST(A)	EACH(B)	OF	THE	GAMES(C)	HE	PLAYED
		11			3			3
		(4)			(2)			(8)
								(4)

The phonetic maxima are spread over several lexical items in sentence 12 and the results *reflect* this spread of phonetic signals, i.e. there is a much wider spread of judgments in 12 compared with 20. In sentence 20 there seems to be a choice between the two elements lost and games, the former containing the cumulation (ABC) and the latter being treated as the rightmost lexical item. In sentence 12, however, each stressed item is selected as the tonic element by several judges. This might seem reasonable since each of the first three stressed elements (lost, each and games) are realised on a phonetic maximum, but the fourth stressed element (played) seems to be being treated as the rightmost lexical item in sentence 12 whereas it is treated as 'given' in sentence 20. There is a further anomaly contained in the respective judgments of the item lost in sentence 12 and in sentence 20. In sentence 20, where it is realised on the cumulation (ABC) lost only receives 4 votes as the single tonic element, but in

sentence 12 where it is realised on only one phonetic maximum (maximum pitch height) lost receives 11 votes as the single tonic element. I can think of no reason for this difference, except perhaps to suggest that when there are two prominent elements, the second is heard as perceptually more prominent than the first, but when all of the stressed elements in an utterance are judged to be equally prominent, then the first stressed syllable will be perceived as the most prominent. This explanation seems to me to be intuitively odd, but I can think of no other reason which might account for these anomalous results.

The results for all seven sentences containing the cumulation (ABC) have been examined, and some sentences containing the same lexical content but different phonetic signals have been compared with the sentences containing the cumulation (ABC). The results have shown that the cumulation (ABC) is not automatically identified as the tonic element. The last lexical item within a unit also seems to be predisposed to be identified as the tonic. It was not the case, however, that the rightmost lexical item was consistently identified as the tonic element, nor was it the case that the cumulation (ABC) was consistently identified as the tonic element. The two 'signals' seem to compete such that the judgments are divided between the two.

5.1.3 Sentences where the PM are spread over several lexical items

Let us now examine sentences which do not contain the cumulation (ABC), but which have the phonetic maxima spread over the lexical items in different ways. Let us compare sentences 4 and 6:

- | | | | | |
|----|-------------|----|--------|-------|
| 4. | THERE (ABC) | IS | MY | HOUSE |
| | 12 | | | 14 |
| | (3) | | | (3) |
| | | | | |
| 6. | THERE (AB) | IS | MY (C) | HOUSE |
| | 2 | | | 27 |

Sentence 4 has already been discussed in section 5.1.2. This sentence is a very good illustration of the equal division of votes between the element realised on the cumulation (ABC) and the rightmost lexical item. If we compare this sentence with sentence 6, we find that when the phonetic maxima are divided between different elements, the rightmost lexical item becomes a 'stronger' candidate and in fact receives the vast majority of votes as the tonic element (i.e. 27 out of 29). Let us examine other sentences where the phonetic maxima are spread over different elements:

- | | | | | | | | |
|----|-------------|---------|-------|----|-----|-----|------------|
| 7. | THE OLD (A) | MAN (C) | ASKED | IF | SHE | HAD | POSTED (B) |
| | | | 2 | | | | 7 |
| | (3) | | (2) | | | | (5) |
| | | | | | | | |
| | THE LETTER | | | | | | |
| | 15 | | | | | | |
| | (4) | | | | | | |

8. HE PLAYED (AC) SQUASH (B) ON MONDAYS WEDNESDAYS
 1 4 10
 (1) (6) (3)

AND FRIDAYS

6
 (4)

In both of the above sentences the majority of judges who selected a single tonic selected the right-most lexical item or constituent (since in sentence 8, 10 judges selected the whole constituent Mondays, Wednesdays, and Fridays). The parameter which seems to attract the largest number of votes is the phonetic parameter of maximum pitch movement (on posted in sentence 7, and on squash in sentence 8). If we compare 8 with sentence 13 discussed at length in section 5.1.2, we find that played is not selected as the tonic by as many judges in sentence 8 when it is not realised on the cumulation (ABC). Otherwise the spread of votes is very similar between the two different realisations of the same sentence.

Let us now examine three different phonetic realisations of sentences containing very similar lexical content:

11. HOW MANY (AC) CHILDREN (B) HAVE YOU THOUGH
 1 27

14. HOW (AB) MANY CHILDREN (BC) HAVE (B) YOU THOUGH
 4 2 16 2
 (2) (1) (4) (1) (1) (1)

17. HOW (A) MANY (C) CHILDREN (B) DO YOU HAVE
27 1 1

In sentences 11 and 17 the element children is selected as the tonic by 27 judges out of 29. In both cases the lexical item children is realised with the maximum pitch movement of the sentence; in sentence 11, the maximum pitch movement is in competition with the combination of maximum pitch height and maximum intensity, and in sentence 17, maximum pitch movement is in competition with each of the other phonetic maxima separately. In both sentences the element realised with maximum pitch movement is selected by the majority of the judges as the tonic. Sentence 14 has an equivalent amount of maximal pitch movement on three elements, how, children and have. As might have been expected the votes are divided, judges showing a preference for the element which combines maximum pitch movement with maximum intensity. I think it would be wrong however to conclude from this example that the combination (BC) acts as a stronger cue for the identification of tonic than either (AB) or (B). The lexical content of this utterance seems to me to play a very important part in determining the 'options' or the possible elements which might be selected as tonic. This sentence contains only one open class substantive, children, the other words are closed class items. Therefore it might be suggested that since this

sentence contains only one 'lexical' item, then that item will be selected as the tonic by the majority. I suspect that sentence 14 has such a wide variety of judgments because the contour sounds auditorily like a straight fall from the highest point how. The element children therefore does not seem to have as much prominence because it is not realised on a peak which 'stands out' from a baseline, as is the case in sentences 11 and 17. In sentence 14 each word in the sentence is selected on at least one occasion as the most prominent element in the sentence. This seems to indicate the autonomy of a contour which is realised as one continuous movement. Despite this 'autonomy' (or perhaps because of it), the open class substantive is agreed by the majority to be the most prominent item in the sentence.

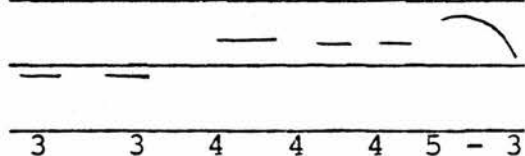
Thus we see from the above sentences that when the phonetic maxima are spread over several items in the sentence, the judgments are also spread over several items in the sentence although the rightmost lexical item seems to be judged by the majority to be the realisation of tonic. Maximum pitch movement also attracts some of the votes but we must also take account of the lexical content of the sentence when accounting for tonic judgments.

5.1.4 Sentences where the rightmost lexical item
combines with phonetic maxima

So far we have been examining sentences where the rightmost lexical item has not been realised with any of the phonetic maxima, let us now examine sentences with such a combination, together with their contour transcription:

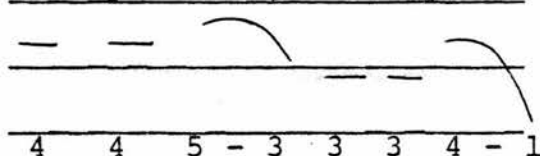
1. HAVE YOU CHANGE (C) OF A POUND (AB)
 3 19
 (7) (7)

have you change of a pound



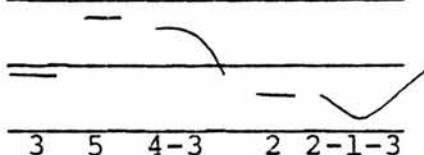
3. HAVE YOU CHANGE (AC) OF A POUND (B)
 5 20
 (4) (4)

have you change of a pound



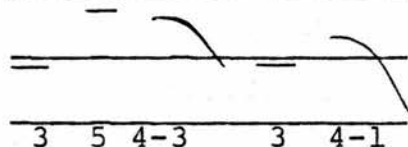
10. MAY (C) I (A) WATCH THE GAME (B)
 4 1 3 20
 (1) (1) (1)

may I watch the game



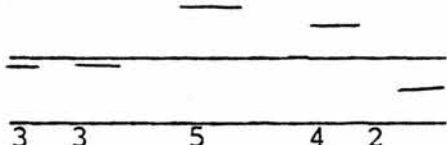
15. MAY I(A) WATCH THE GAME(BC)
 10 4 12
 (3) (3)

may I watch the game



18. I HAVE THREE(AC) DAUGHTERS(BC)
 7 18
 (1) (3) (4)

I have three daughters



As we can see from the transcriptions, each of these sentences consists of a single-peaked contour (excluding perhaps sentence 15) and the phonetic maxima are spread over different parts of the contour. In all of the above sentences the rightmost lexical item combines with the maximum movement of the sentence. Both Halliday and Crystal state that the tonic is always realised on the rightmost lexical item in an unmarked sentence, and that the tonic is identified as being the most prominent item or the item containing the most pitch movement.

This experiment was set up primarily to find out whether 'prominence' or 'tonic' was consistently identified by a cumulation of three phonetic maxima. We found

that this was not the case. The second task

was to see which phonetic maximum of the three under consideration would be recognised as the cue for tonic placement when the cumulation of maxima was not present. The results indicate that a single phonetic maximum alone is not necessarily a sufficient cue for tonic identification, since although maximum pitch movement seemed to be a strong cue, in sentences 7, 8, 12, the rightmost lexical item was preferred over the element containing the maximum pitch movement.

In sentences 1, 3, 10, 15 and 18, the two strongest cues for tonic identification combine on one element since in all of these sentences the rightmost lexical item is realised with the maximum pitch movement of the sentence. It is not the case, however, that every judge selects the rightmost lexical item as the tonic element.

5.1.5 Conclusion

We can conclude that there are three very strong cues for tonic identification, these are:

- 1) the cumulation of phonetic maxima (ABC)
- 2) the rightmost lexical item
- 3) the maximum pitch movement

It is not the case, however, that all of these cues combine in every utterance. Nor is it true that the

presence of one of these cues (in the absence of the other) will automatically determine the tonic element. On the contrary all of these cues may compete within a single utterance and each of the cues may be overridden depending on, for example, the information structure of the utterance (cf. sentence 20 where played is treated as given) or the composition of the utterance (cf. sentences 11, 14, and 17 which contained only one open class substantive). Tonic is thus seen to be a complex phenomenon which cannot be simply defined either in phonetic terms or in terms of clause / sentence structure.

5.1.6 Judge groupings

The judges taking part in this experiment were selected from different accent areas in order to test the hypothesis that the phonetic cues of one dialectal area may not be interpreted in the same way by speakers from other areas. The judges were therefore grouped in the following way:

<u>NATIVE ENGLISH SPEAKERS</u>		<u>NON-NATIVE ENGLISH SPEAKERS</u>	
English English	9	Arabic	3
Welsh English	1	Icelandic	2
Scottish English	2	German	1
South African English	1	Spanish	1
Australian English	1	Thai	1
American English	6	Malay	1

Non-native speakers were included in the experiment in order to see whether they would be more likely to listen for specific phonetic cues, or whether they would select the rightmost lexical item as suggested by Halliday. All of the judges were familiar with Halliday's definition of 'tonic'. The various groups were compared to see if there was any significant preference for cues according to language background.

It was interesting to note that all of the native English speakers tended to adopt a strategy of some kind which was very consistent throughout the experiment. The two Icelanders and two of the native Arabic speakers also adopted a strategy which they used fairly consistently throughout the experiment. The remaining 5 non-native speakers did not adopt a main strategy, but selected tonics according to various realisations of phonetic cues as well as according to rightmost lexical item. No consistent strategy was adopted by this group. 14 out of 29 judges chose movement and rightmost lexical item as their main cues for tonic identification. This group consisted of 8 English English speakers, the Welsh English speaker, one Scottish English speaker, the South African, the Australian, one American and one of the native Arabic speakers.

7 judges, consisting of the 2 Icelanders, one Scottish English speaker, and 4 American speakers, selected tonics according to the parameter of maximum height. The

last group of three judges (1 American, 1 English English speaker, and one native Arabic speaker) selected tonics according to the rightmost lexical item, not really paying much attention to the phonetic parameters at all.

It seems to be the case, therefore, that individuals do vary according to the strategies adopted as cues for tonic identification. The accent / language groups tested in this experiment are not large enough to allow any conclusions to be drawn about the effect of language background on these strategies, but there seemed to be a definite tendency for English English judges to prefer the cues of pitch movement and rightmost lexical item, whereas the American judges seemed to prefer the cue of maximum pitch height.

The results were also examined according to the phonetic experience of the judges, the hypothesis being that an experienced phonetician would be able to hear all of the phonetic cues thus being unable to choose between them if they occur in competition, whereas a student who has just been introduced to the notion 'tonic' may well adopt the strategy of rightmost lexical item as the cue for tonic identification according to Halliday's or Crystal's definition.

The judges were grouped according to the amount of experience in phonetics each subject had. These groupings were as follows:

A. Professional phoneticians with several years teaching experience	4
B. PhD students in phonetics	5
C. PhD students in general linguistics	2
D. Linguists with some experience in teach- ing phonetics	4
E MSc. students, post-graduate language teach- ers, recently introduced to the notion 'tonic'	14

There were few neat generalisations arising from the above groupings the only real generalisation being in the form of 'the more you know, the more you know you don't know'. In other words the results conformed to the hypothesis that the group of phoneticians with most experience in phonetics found it very difficult to make decisions about a single tonic placement. This group tended to identify any lexical item which was realised on any phonetic maximum or combination of maxima as a tonic element. The results of this group indicate that if we are to divide up a sentence into tone units, with each unit containing one tonic, then the tone unit in many cases will contain only one stressed syllable and would look very like the unit of rhythm referred to as the 'foot' (cf. sentences 1, 3, 7, 11, 12).

There was only one group who behaved consistently as a group, and that was group D. The members of this

group had taught basic phonetics, but were primarily concerned with other areas of linguistics. All of the judges in this group were native speakers of English (consisting of 1 American, 1 Scot, and 2 Englishmen). Perhaps this group had just enough phonetic experience to perceive phonetic prominence, but not too much experience that would enable them to hear all of the phonetic cues present.

All of the members of the remaining groups behaved according to their language backgrounds. For example in group A, which consisted of 2 American English speakers and 2 English English speakers, the two Americans selected tonics according to pitch height, whereas the two English English speakers selected tonics according to pitch movement and rightmost lexical item.

5.1.7 Overall results

The overall results, therefore, indicate that there are several main parameters which function as cues for the identification of the tonic element of an utterance. These are:

- 1) the cumulation of phonetic maxima (ABC)
- 2) the rightmost lexical item
- 3) maximum pitch movement
- 4) maximum pitch height

These cues may contradict each other, or they may conflict with other cues involving syntactic and semantic criteria. The background of hearers who are performing the task of identifying 'tonic' must also be taken into consideration, since it seems to be the case that American English speakers tend to prefer pitch height as a cue, whereas English English speakers tend to prefer pitch movement as a cue for tonic identification. This tendency is reflected in the two 'schools' of information theory: the one stemming from America being based on pitch levels or heights; the one stemming from Britain being based on the direction of pitch movement (cf chapter 8).

Since different judges select different elements as the tonic for different reasons, and since judges will select different elements as equally prominent, again for different reasons, the notion 'tonic' seems to be a very complex one. This experiment was therefore followed by a second experiment which tried to restrict some of the syntactic and semantic criteria involved in the task of tonic identification.

5.2 Experiment 2

5.2.0 Introduction and Hypotheses

Halliday (1967) maintains that the clause is the domain for various systems; the system of 'information', the system of 'thematization', and the system of 'identi-

fication'. The system of 'information' is realised by phonological features of intonation, the system of 'thematisation' is realised by the sequence of elements in the clause, and the system of 'identification' is realised by specific patterns of clause structure.

In Halliday's terms one information unit is realised as one tone group. Each information unit has either one primary point of information focus, or one primary followed by one secondary (realised by tones 13 or 53). The focal element or the focus of information will occur on 'new' information, or information which is not recoverable from the discourse. In the unmarked case, the focus of information falls on the final element of the information unit (note that the rightmost lexical item in experiment 1 was a strong candidate for tonic, but was not selected as the tonic in every instance).

Contrastive sentences are treated by Chafe (1976) as a 'special case' of new information. Chafe states (p. 35) that contrastiveness is realised by "the placement of higher pitch and stronger stress on the focus of contrast. Thus it is often difficult or impossible to tell the difference between contrast and new information on a phonetic basis alone." But Chafe continues by saying (p. 36) "On the other hand it is also true that the high pitch on a contrastive focus is often higher than that on a simple new information focus. Sentences which contain an element which specifically

contradicts a previous element will be said to contain an element in lexical contrast where one lexical item can be seen to contradict another, e.g. given the conversation between A and B as follows:

A: Did John wash the car?

B: No.

C: Did Merlin wash it then?

Merlin is a realisation of lexical contrast.

Hypothesis: The element in lexical contrast will be identified consistently as the tonic element when the utterance containing lexical contrast is presented out of context.

Halliday states that the clause in his system of thematisation is organised into the structure 'theme' and 'rheme'. Initial position in the clause is assigned the status of theme, and the rheme is all that follows. The theme may be isolated as a separate information unit especially if the clause is structured in the identifying equative form (i.e. if the clause is a cleft structure). Thus Halliday notes that both the identifier and the identified can carry information focus (p. 226), or the identifying clause may be realised as a single inform-

ation unit with only one of the two elements focal. If this is the case, then the single focus normally falls on the identifier irrespective of sequence. Thus in the following examples cited in Halliday 1967, p. 226, the underlined elements could carry the information focus:

- a) //the one who painted the shed last week//
was John//
- b) //John//was the one who painted the shed
last week//
- c) //the one who painted the shed last week
was John//
- d) //John was the one who painted the shed
last week//

((a) and (b) are sentences where both the identifier and the identified are realised with information focus, while (c) and (d) carry only one information focus, in each case realised on the identifier John)

Jespersen in 1961 notes that "A cleaving of a sentence by means of it is (often followed by a relative pronoun or connective) serves to single out one particular element of the sentence and very often, by directing attention to it and bringing it, as it were, into focus, to mark a contrast". (p. 147-8). Chafe (1976) agrees with this view of cleft sentences stating on p. 37, "An ad-

ditional way in which contrastiveness may be expressed is with the use of a so-called cleft sentence, as in It was Ronald who made the hamburgers."

Cleft sentences were also included in experiment 2 since they also express contrastiveness (or focus of contrast) but by means of word order.

Hypothesis: The identifier of a cleft construction will be selected consistently as the tonic element. The identifier may be realized as the single tonic of the cleft construction, or it may be one of two, the second tonic occurring on the identified element of the construction.

5.2.1 Method

In order to obtain cleft sentences and contrastive sentences, a game was set up involving two players (P1 and P2). P1 was given a story to read (see Appendix 5) prior to recording, and P2 was given a sheet of paper containing a list of characters and a list of actions (see Appendix 6). P2 was then asked to match up the characters and the actions, and to attempt to retell the story which by then was familiar to P1. P2 was only permitted to ask 'yes/no' questions, and P1 was asked to answer 'yes' or 'no' unless P2 was desperately in need of help, when no more clues could be given. It was thought

that in this situation we would obtain several types of questions, including cleft questions, and that contrastives would be produced in this test which involved 'filling-in-the-slot' procedures, i.e. since P2 was forced into producing 'yes/no' questions, he would have to specify both the character and the action which he was attempting to match, thus being forced into either constructions of the type 'Did X do Y', or constructions of the type 'Was it X that did Y'. If the attempted matching failed, he might then produce a contrastive construction such as 'Did X do Y', 'no', 'Well did Z do Y'. These constructions were in fact obtained:

R2/2 A: Was it the rich farmer who had three sons...

(an example of a cleft sentence)

L1/1 A: Well did the old man have three sons?

B: mm, no.

A: Did the rich farmer have three sons...

(where the underlined element is contrastive)

20 sentences were selected from three sets of games. 18 of the sentences were taken from the recordings of two male Scots speakers, and two of the sentences were taken from one female English speaker. Of these 20 sentences, 11 are examples of cleft sentences, and in all 11 sentences the clefted element is the identifier of the sen-

tence. 3 of these clefted sentences contain a clefted element which is also in lexical contrast. 7 sentences are non-cleft contrastive sentences, and 2 are non-cleft non-contrastive sentences.

Judges were asked to identify the 'tonic of the sentence'. If they felt there was more than one tonic, they were asked if possible to rank them in order of importance.

5.2.2 Results: Sentences containing the cumulation (ABC)

17 out of the 20 sentences in experiment 2 contained an element with the cumulation (ABC) (see Appendices 7 and 8 for full details of the results). These sentences included all seven contrastive sentences where the cumulation (ABC) was realised on the contrasted element; 8 cleft sentences where the cumulation (ABC) was realised on the clefted element, i.e. the identifier; and the 2 simple declarative sentences where the cumulation was realised on the leftmost lexical item. Of the 17 elements containing the cumulation (ABC), only 12 are judged to be the tonic element by more than 20 out of the 25 judges who took part in the experiment (5 of the 11 clefted elements; and all 7 contrasted elements).

All of the sentences containing elements in lexical contrast had the cumulation of phonetic maxima realised on the contrasted element. All contrasted elements were

judged to be the tonics of the sentences involved by at least 23 out of 25 judges. A contrasted focus of information, therefore, seems to be identified consistently as the tonic element. These sentences are given below, with the contrasted element underlined:

1. THE PONY(ABC) RUNS AWAY
25
6. DID THE RICH FARMER(ABC) HAVE THREE SONS
23 0
(2) (2)
8. THE DOG(ABC) RUNS AWAY
23 2
16. IS IT THE MILLER'S(ABC) DOG
25
3. PRESUMABLY IT WAS THE ELDEST(ABC) SON
25
11. IT WAS THE ELDEST(ABC) SON
25
20. SO THEY GOT(ABC) MARRIED
1 24

The two simple declarative sentences which contain the cumulation (ABC) which is not identified consistently as the tonic are as follows:

4. THE SECOND(ABC) SON MARRIED THE MILLER'S DAUGHTER(AB)
2 12
(11) (11)
17. WELL DID THE OLD MAN(ABC) HAVE THREE SONS(AB)
0 5 7
(2) (11) (11)

In sentence 4 the cumulation of phonetic maxima falls on second son, but there is also maximum pitch movement and maximum pitch height on millers' daughter. The phonetic maxima are therefore spread over both elements. If we examine the preceding lexis, we find that both elements are being introduced for the first time in this utterance, thus both elements are 'new'. 11 out of the 25 judges select both elements as tonic, thus electing to assign a unit of information to each 'new' element as predicted by Halliday. 14 judges select only one tonic element in sentence 4, 12 of these selecting the element which is the rightmost lexical item realised with maximum pitch movement, the remaining 2 selecting the cumulation (ABC) as tonic. The same confusion of cues which was noted in section 5.1.7 applies to the above sentences. In sentence 17, the cumulation (ABC) falls on old man, but there is also maximum pitch movement and maximum pitch height on three sons, and both constituents are being mentioned for the second or third time, therefore both constituents are treated by the speaker as 'given'. In sentences 4 and 17, therefore, there is the same confusion of cues in operation which was noted in section 5.1. Both elements within the sentence have the same status on the given / new dimension, and the phonetic cues are ambiguous as to which of the two should be preferred, the logical answer then should be for the judges to select both elements consistently

as two realisations of tonic. 11 out of 25 judges do exactly this in both sentences, but even more judges select only one tonic in the above sentences, and are divided between the two elements which are 'marked' by various cues as potential tonics. There is therefore substantial disagreement between the judges as to whether these sentences contain a single tonic or two tonics. There is also disagreement as to where the single tonic occurs.

Let us now examine the cleft sentences. Only 5 out of 8 sentences which contained the cumulation (ABC) on the clefted element had the clefted element selected as the tonic. Three of the five are those cleft sentences whose clefted element is also an element in lexical contrast. These sentences therefore support the conclusion that contrastive focus is consistently identified as tonic. These sentences are given below:

10. WAS IT THE RICH FARMER'S ELDEST SON (ABC)

23
(2)

WHO MARRIED THE GENTLEMAN'S DAUGHTER

0
(2)

13. WAS IT THE MILLER'S DAUGHTER (ABC) WHO WAS

23
(2)

VERY BEAUTIFUL THEN

(2)

15. WAS IT THE BAD-NATURED DAUGHTER (ABC)

21
(2)

WHO RAN AWAY

2
(2)

(where the element in lexical contrast is underlined).

The remaining sentences which had the cumulation (ABC) on the clefted element, contain 'new' identifiers which are not recoverable from the preceding discourse, while the identified element of the sentence has been 'given' immediately previously. Thus since the identified element has been repeated in these sentences after having been mentioned immediately previously, there is an unequivocal information focus on the 'new' identifier, or clefted element, which is judged to be the tonic element by more than 80% of the judges (i.e. more than 20 out of 25). These sentences are given below:

9. WAS IT ONE OF THE SONS (ABC) WHO WAS VERY WELL OFF
23 0
(2) (2)

12. WAS IT THE YOUNGEST (ABC) SON THAT MARRIED THE
20 0
(5) (1)

BEAUTIFUL MILLER'S DAUGHTER
0
(4)

Therefore, if contrastive sentences are seen to be a special case of information focus as suggested in section

5.2.1, then it is clearly the case that an unequivocal information focus is identified consistently as the tonic element.

Let us now examine the remaining cleft sentences. Three cleft sentences contain the cumulation (ABC) on the clefted element which is not identified consistently as the tonic element. These sentences are as follows:

2. WAS IT THE MILLER'S DAUGHTER (ABC) WHO RETURNS
9
(7)

HOME ON THE BACK OF THE PONY
0 9
(3) (7)

5. WAS IT THE YOUNGEST SON (ABC) WHO TELLS HIS

8
(11)

WIFE TO CARRY THE SADDLE
6
(11)

19. IS IT THE PONY (ABC) WHO REFUSES TO CROSS A
11
(8)

STREAM AND IS SHOT (AB)	
0	5
(3)	(8)

These sentences are spoken by the same speaker, and are very similar in length and in number of constituents. In sentence 2 milller's daughter has just been mentioned in the previous sentence and the remaining constituents are mentioned for the first time. In sentence 5 youngest son has been mentioned only two sentences previously and

the remaining constituents are being mentioned for the first time. In sentence 18 pony has been mentioned previously (about 16 sentences before) and the remaining constituents are being mentioned for the first time. Therefore in all of the above sentences, the clefted element or identifier is 'given', and the identified element is 'new'. According to Halliday, when there is one focus of information in a cleft construction, the focus will be realised on the identifier (in these sentences the clefted elements), yet the focus of information is also said to be realised on the 'new' element of an utterance (in the above sentences on the 'identified element' which is realised on the last lexical item -- a strong cue as noted in 5.1). There is therefore a conflict here which is reflected by the tonic judgments. Let us examine the acoustic records of elements which are selected as tonic elements.

In sentence 2, millers daughter measures 150-220-130cps, whereas pony measures 200-140cps. In sentence 5, youngest son measures 140-210-140cps whereas saddle measures 190-130cps. In sentence 18 pony measures 210-135cps whereas stream measures 185-145cps, and shot measures 200-130cps. The maximum intensity occurs on the clefted element in all three sentences, and decreases considerably towards the end of the sentence. The conflict in the information structure is therefore reflected in the

phonetic signal.

If each element which has been judged as a tonic in these sentences was the focus of information of separate information units then each judge should have selected all of the elements as realisations of tonic, but this was not the case, e.g. in sentence 2, 18 judges decided that there was only one element functioning as the tonic of this sentence, 9 out of 18 selected millers daughter which is the identifier and contains all 3 maximum phonetic values; 9 out of 18 selected pony as the tonic element, which is the 'new' element in the information structure and is the rightmost lexical item in the sentence. The remaining 7 judges selected both of these elements as equivalent tonics. These results do not suggest that there are two units of information, they suggest rather that there is a conflict of cues for tonic identification which is not resolved by the sentence structure.

The three remaining cleft sentences do not contain the cumulation of phonetic maxima on the clefted element, cf. below:

7. WAS IT THE SECOND SON (AC) WHO MARRIED THE MILLER'S
 12
 (7)

DAUGHTER (B)
 6
 (7)

14. WAS IT THE WIFE (AC) OF THE NEWLY-MARRIED COUPLE (B)
 6 5
 (12) (13)

14. ...WHO WAS IMPRESSED (B) BY PREVIOUS EVENTS (A)

0
(3)

0
(4)

AND BECAME A VERY OBEDIENT WIFE

0
(1)

19. WAS IT THE RICH FARMER (Bc) WHO HAD THREE SONS (A)

10
(7)

8
(7)

Sentence 19 is very similar to the sentences we have just been discussing. The clefted element has just been mentioned previously and the remaining elements are being introduced for the first time. The only difference between sentence 19 and the previous group of sentences is that the clefted element in 19 is not realised on the cumulation (ABC). The phonetic maxima are divided between two elements.

Sentence 7 also has a spread of phonetic maxima, and all of the elements are being introduced for the first time, there is no unequivocal instance of 'new' versus 'given'. Sentence 14 is very long, and was introduced to see whether the clefted element would be preferred as the tonic when there was such a variety of phonetic cues. The clefted element is in fact 'preferred', although the judges are undecided as to which part of the clefted element should be the tonic. All of the elements containing any phonetic maxima are selected as tonics by various judges. Sentences

1, 5, 7, 14, 18, and 19 do not contain an unequivocal instance of information focus either according to the information structure of the preceding discourse or according to the placement of 'tonics' in these sentences. All of these sentences are cleft sentences where the clefted element is the identifier. Halliday would predict, therefore, that the clefted element should be selected as the tonic element if only one tonic is selected, but this is not the case. Even when the identifier is realised on the cumulation of phonetic maxima, it is not consistently identified as the tonic. The cues of rightmost lexical item conflict with the cues of clause structure, and the cues of phonetic maxima. It is not enough to account for the above results by dividing the sentences into two information units, since if there had been two obvious information units, the majority of the judges should have selected two information foci, but they did not. Some judges selected tonic according to rightmost lexical item; others according to cumulation of phonetic maxima; and still others according to clause structure. Judges were not always consistent in their choice of cues.

5.2.3 Conclusion

Despite the fact that this experiment attempted to control the sentences so that the focus of information would be predictable, the results show:

1) where the information focus is realised on a contrasted element, or a 'new' element which is not recoverable from the preceding discourse, then the information focus is consistently identified as the tonic of the utterance.

2) where there are two or more elements within an utterance which have the same status according to the preceding discourse (i.e. both elements are 'new' or both elements are 'given'), both or all of these elements will compete as tonics of the utterance. The phonetic cues, and the cues of clause structure which act as cues for tonic identification conflict so that either element, or both elements may be identified as tonic.

Experiments 3, 4 and 5 were set up in order to examine the effect of the discourse on tonic identification.

5.3 Experiments 3, 4 and 5

5.3.0 Method

The two previous experiments consisted almost entirely of Scots data (apart from 2 sentences in experiment 2 which were taken from the recording of an English speaker). Experiments 3, 4 and 5 contain only RP data to see whether Halliday's predictions about tonic identification would be more consistent for RP data than for Scots. The data for these experiments was again taken from recordings of the

game situation outlined in 4.1.6 and in 5.2.1.

Experiment 3 consisted of 40 sentences taken out of context and presented to the judges on tape. The judges also had a written transcription of the sentences in front of them. Experiment 4 consisted of 17 of the previous 40 sentences presented in the context of the surrounding discourse in written form only. Experiment 5 consisted of exactly the same material as experiment 4, but the material was presented auditorily on tape, together with the written transcription.

Different numbers of judges took part in each experiment. 10 judges took part in all three experiments (3, 4 and 5) which were presented in the order 3-4-5, with several weeks in between each experiment. The total numbers of judges taking part in each experiment were as follows:

22 judges took part in experiment 3

10 judges took part in experiment 4

16 judges took part in experiment 5

All of the judges were familiar with Halliday's notion of tonic, the range of experience in phonetics being similar to that stated in 5.1.6. The 10 judges who took part in experiments 3, 4 and 5 had also taken part in experiments 1 and 2 at an earlier date. There-

fore there was a consistent core of judges throughout the series of experiments. The judges were again asked to select a single tonic if possible and if they wished to identify 2 or more tonics to try to rank them in order of importance.

5.3.1 Hypothesis

Experiment 2 had suggested that the preceding discourse structure affected tonic identification so that a 'contrasted' element was selected consistently as tonic despite the fact that the judges did not have access to the preceding discourse.

Experiments 3, 4 and 5 were therefore set up with the hypothesis that 'tonic' would be identified as occurring on the same element(s) when judges were presented with the same utterance (a) out of context with phonetic cues; (b) in context without phonetic cues; and (c) in context with phonetic cues.

5.3.2 Results: experiment 3

Let us examine the results of experiment 3 briefly before going on to examine and compare the results of all three experiments together.

The sentences in experiment 3 can be divided broadly into three groups according to the information structure of the preceding discourse. The first group include those sentences which contain an element in lexical con-

trast, i.e. an element which specifically contradicts another element in the preceding utterance. These sentences are as follows (with elements in lexical contrast underlined):

A) sentences containing lexical contrast

3. ITS NOT THE YOUNGEST SON(ABC) WHO DOES IT
 0 19
 (3) (3)

4. WAS THE RICH(AC) FARMER(ABC) VERY WELL OFF
 1 15 0
 (6) (5) (1)

5. IS IT THE MILLER'S(ABC) DAUGHTER WHO IS
 15 0
 (7) (1)

VERY BEAUTIFUL

 0 0
 (2) (4)

8. THEN IT WAS THE YOUNGEST(ABC) SON WHO DECIDED
 17
 (5)

TO GET MARRIED

 0
 (5)

9. IS IT THE MILLER'S(ABC) DAUGHTER WHO WAS
 18 0
 (4) (1)

VERY BEAUTIFUL

 0
 (4)

10. THE DOG(ABC) RUNS AWAY
 22

24. WAS IT THE RICH FARMER'S(ABC) ELDEST SON WHO
 0 0 17 1
 (2) (2) (3)

MARRIED THE GENTLEMAN'S DAUGHTER

 0
 (1)

32. DID THE PONY (ABC) REFUSE TO CROSS THE STREAM
 15 0 0 0
 (7) (1) (1) (5)

34. WAS THE GENTLEMAN'S (ABC) DAUGHTER VERY
 20
 (2)

BEAUTIFUL
 0
 (2)

35. DID THE ELDEST (ABC) SON MARRY A MILLER'S
 18 0
 (4) (2)

DAUGHTER
 0
 (2)

All of the above sentences contain one element of lexical contrast which is realised on the cumulation (ABC). All of these elements are identified as tonics by all of the judges. When judges select more than one tonic in the sentence the contrasted element is judged to be the most important tonic in the sentence.

The next group of sentences consists of those sentences which contain a 'new' element, and the rest of the sentence is 'given'. This group can be sub-divided into two sub-groups according to whether the 'new' element is identified consistently as the tonic element or not. Let us first examine those sentences where the majority of judges identify the 'new' element as tonic (with 'new' element underlined):

B) Sentences containing a 'new' element

13. SO DID(ABC) THE YOUNGEST SON MARRY THE BAD-
 14 0 0 0
 (8) (1) (1) (1)

NATURED DAUGHTER

$$\begin{array}{cc} 0 & 0 \\ (3) & (4) \end{array}$$

15. WAS (C) IT THE BAD (B) - NATURED (B) DAUGHTER (ABC)
- | | | |
|-----|-----|-----|
| 0 | 0 | 17 |
| (2) | (1) | (5) |

WHO RAN AWAY

0
(4)

21. IS (A) IT THE RICH FARMER(ABC) WHO HAS A BAD-
0 11
(3) (9)

NATURED WIFE AND A DAUGHTER WHO IS WORSE

$$\begin{array}{cccc} 0 & 0 & 0 & 0 \\ (3) & (6) & (4) & (6) \end{array}$$

22. IS (C) THE PARENT (ABC) OF THE BAD-NATURED
- | | | |
|-----|-----|-----|
| 0 | 12 | 0 |
| (1) | (9) | (2) |

DAUGHTER VERY WELL OFF

0 1
(2) (5)

29. WAS IT THE SECOND (ABC) SON WHO MARRIED A
- | | | |
|------|-----|-----|
| 11 | 0 | 0 |
| (10) | (1) | (2) |

MILLER'S DAUGHTER

0 1
(1) (7)

36. DOES THE OLD MAN (ABC) HAVE SONS AS WELL (B)
- 0 2 0 1 11
- (1) (4) (4) (8)

37. IS IT THE OLD MAN(ABC) WHO WAS VERY WELL OFF
- | | |
|-----|-----|
| 21 | 0 |
| (1) | (1) |

38. THE YOUNGEST SON (ABC) MEETS AN OLD MAN
- | | |
|-----|-----|
| 15 | 0 |
| (7) | (7) |

40.	OH	HE'S	THE	ONE	WHO	HAS	(ABC)	A	BAD-NATURED	WIFE
	0		0			15			0	
	(4)		(1)			(6)			(1)	

All of the above sentences have the 'new' element identified as the tonic element by at least 20 out of the 22 judges. When the judges select more than one tonic in the sentence, the new element is judged to be the most important tonic. All of the new elements are realised on the cumulation (ABC) except for sons as well in sentence 36. This element is realised with maximum pitch movement and constitutes the rightmost lexical constituent. This combination was found to be a very strong cue for tonic identification in experiment 1, equal to the cue of the cumulation (ABC).

If we examine the above group of sentences we find that 6 of these 9 sentences have the new element realised on the leftmost lexical item of the sentence, and all six are realised by the cumulation (ABC). In sentence 13, the cumulation (ABC) is realised on the element did. The speaker is questioning the relationship between the youngest son and the bad-natured daughter, therefore did may also be regarded as a new element realised as the leftmost lexical element of the sentence. Sentences 36 and 40 are slightly more complex since each of these sentences contains an additional phonetic maximum as well as the cumulation (ABC). The presence of this additional phonetic cue to prominence means that the votes

are divided between the two phonetic signals, with the result that more judges select more than one tonic per utterance compared to the results obtained for the contrastive sentences.

There are other sentences containing a 'new' element which is not identified consistently as the tonic. These sentences are as follows, with the new element underlined:

1. SO ITS (ABC) THE BAD-NATURED (B) DAUGHTER WHO

2
(11)

2
(14)

DECIDED TO GET MARRIED

0
(12)
19. WAS THE OLD (A) MAN (BC) VERY WELL OFF

5
(6)

3
(3)

8
(5)
27. SO IT WOULD BE THE RICH FARMER (ABC) WHO HAD

0
(1)

0
(5)

7
(12)

THREE SONS

0
(4)

0
(6)
33. WELL I PRESUME (BC) THAT IT WAS THE BAD-NATURED

0
(3)

0
(14)

0
(1)

(5)

DAUGHTER (ABC) WHO WAS BEAUTIFUL

7
(11)
39. IS THE BAD-NATURED DAUGHTER (ABC) VERY BEAUTIFUL

1
(7)

1
(3)

12
(7)

All of the above sentences (apart from sentence 33) have the new element realised as the rightmost lexical

element of the sentence, yet none of these are selected by the majority of judges as the tonic element. Sentence 33 has the cumulation (ABC) realised on the clefted element, bad-natured daughter, which competes with the leftmost lexical element presume, realised as (BC). Both elements are identified by 14 out of the 22 judges as equal tonics. All of the other sentences in this group contain a new element on the rightmost lexical item which is not realised on the cumulation (ABC). Instead, the new element competes with other elements in the sentence (usually the leftmost lexical constituent) which is realised with the cumulation (ABC).

When a new element is realised on the rightmost lexical item, judges tend to select more than one tonic, selecting as tonics elements which contain phonetic maxima. When the new element is realised on the leftmost lexical item, this element is realised by the cumulation (ABC) and is normally selected as the single tonic element of the sentence. It is not true to say therefore, that a new element will be identified consistently as the tonic element of the utterance in which it occurs.

The last group of sentences which might be said to contain new elements are the following, again with the new element underlined:

16. DID (ABC) THE DOG (ABC) REFUSE TO CROSS A STREAM (B)
 0 15 0
 (1) (7) (7)
17. THE HORSE (ABC) REFUSES (BC) TO CROSS A STREAM
 15 0 0
 (5) (1) (2)
 AND IS SHOT (B)
 2
 (5)
25. THEREFORE IT MUST BE THE SECOND (ABC) SON
 0 0 7
 (2) (1) (15)
 WHO MARRIES A MILLER'S DAUGHTER
 0 0 0
 (8) (4) (2)
26. THE YOUNGEST (ABC) SON DECIDED TO GET MARRIED
 10 0 0 0
 (8) (6) (2) (9)

The above sentences do not contain 'new' elements as opposed to 'given' elements since all of the elements have the same status according to the preceding discourse. Because of the nature of the game situation, however, we can expect the relationship between the elements to be 'new'. The game situation is constructed in such a way that the player who is attempting to reconstruct the story has a limited number of strategies available to him. Sentences 16, 17, 25 and 26 are examples of one of these strategies. The player has set up a 'slot' whereby he decides to keep the action constant while attempting to 'fill in' the correct character who performed the action. Thus in sentence 17, the slot has been set up as follows: 'X refuses to cross a stream and is shot'. The

player has tried to fill in the correct character on a previous occasion, but has failed. This slot has been dropped in favour of another, but the player still has the concept of the former slot established in his mind. Thus although all of the elements in sentence 17 have been discussed previously, when the player introduces horse into the slot 'X refuses to cross a stream and is shot', the relationship is 'new'. The speaker treats the action as given since he has already tried to fill in the leftmost element of this structure, therefore for this speaker, the leftmost element is the 'new' element. The same observations apply to sentences 25 and 26.

Sentence 16 is slightly different. All of the elements in sentence 16 are being introduced for the first time, but since the speaker goes on to say "Did the horse refuse to cross a stream" immediately after sentence 16 after receiving the answer 'no', it is apparent that the speaker has already set up the slot 'X refuses to cross a stream', treating the X element as new, and the remainder of the slot as given.

In all of the above sentences, the speaker has elected to use a particular strategy which has involved treating the action as 'given', and the characters as 'new' elements in the relationship. The results indicate that the speaker's strategy is available to the

hearer since the above sentences have been presented out of context. The majority of the judges (18-22) identify as the tonic element, the element which is treated as new by the speaker. The results conform to the previous results of those sentences which contain a new element realised as the leftmost lexical item of the sentence, i.e. the new element is realised by the cumulation of maxima (ABC) and is identified consistently as the tonic element of the utterance.

C) Sentences containing elements of equal status

All of the following sentences contain elements which are 'equal' according to their status as 'given' or 'new', i.e. either all of the elements in the sentence are being mentioned for the first time in the discourse, or they have all been discussed in the discourse previously. The sentences in this group are as follows:

2. THE DOG(ABC) RUNS AWAY(B)

5	1
(16)	(16)

7. SO THE RICH FARMER(ABC) HAD AN ELDEST SON(ABC)

0	0	1	0
(4)	(3)	(8)	(7)

WHO MARRIED A GENTLEMAN'S DAUGHTER

0	3
(2)	(14)

10. IS THE PERSON(ABC) WHO HAS A BAD-NATURED WIFE

3
(18)

AND A DAUGHTER WHO IS WORSE (B) THE RICH FARMER (B)

1 0
(15) (13)

11. IS IT (AB) THE RICH FARMER (BC) WHO HAD

0 0 1
(1) (3) (14)

THREE SONS (B)

0 3
(3) (15)

12. ALL (AC) THE SONS (BC) MENTIONED ARE SONS (BC)

3 0 0 0 0
(11) (1) (6) (2) (9)

OF THE SAME FATHER

2
(10)

14. WAS THE OLD MAN (AC) VERY WELL OFF (B)

5 3 0 4
(5) (7) (1) (7)

20. IT (A) WAS THE YOUNGEST (BC) SON WHO ASKED TO

5 0 0
(13) (3) (1)

MARRY THE BAD-NATURED DAUGHTER

0 2
(1) (13)

28. IS IT THE DOG (ABC) THAT RUNS AWAY (AB)

4 0 7
(11) (1) (11)

30. THE YOUNGEST (BC) SON DECIDED (B) TO GET MARRIED

4 2 0 0
(10) (6) (8) (8)

31. THE ELDEST (ABC) SON MARRIED A GENTLEMAN'S

1 0 0
(17) (5) (11)

DAUGHTER (B) WHO WAS VERY WELL OFF

0 0 0
(7) (3) (10)

All of the above sentences contain various combinations of phonetic maxima which compete with each other as cues

for tonic identification. The majority of judges (ranging between 11 and 18 per sentence) select at least two tonics per sentence if not more. These results reflect the equal status of the elements according to the preceding information structure, but cause problems for a Hallidayan analysis of tone groups. If these utterances are to be divided into tone groups so that each tone group contains only one tonic, then the domain of the tone group must shrink from Halliday's proposed clause-type length, to something which is constituent-type length. See for example sentence 30, which would have to be divided as follows:

30. The youngest/son/decided/to get married/

If the tone group is realised as such a small unit, the notion of information unit also becomes problematic. For example how are sentences 2, and 12 to be divided in order to reflect the division of the utterance into information units? Which possibility is 'correct', or should tone group boundaries be indefinite?

2. The dog/ runs away or The dog runs / away

12. All the sons / mentioned / are sons / of the
same father

or

All / the sons mentioned are / sons / of the
same father

Let us now examine the three experiments together, in order to see whether the context would resolve such problems raised by this last group of sentences.

5.3.3 Results: experiments 3, 4 and 5

17 of the 40 sentences presented in experiment 3 together with the surrounding discourse constituted experiments 4 and 5 (the sentence numbers of experiment 3 will be retained for the purposes of discussion). Experiment 4 was presented to judges in the form of the text only (see Appendix 9); experiment 5 presented the text together with a tape. Experiment 3, therefore, presented the data out of context in both written and aural form; experiment 4 presented the data in context, in written form only; and experiment 5 presented the data in context in both the written form and the aural form. The results of the 17 sentences which the three experiments had in common can be compared and examined in the light of the different kinds of information available to the judges.

The sentences in experiments 3, 4 and 5 can be divided into the same kinds of groups discussed in section 5.3.2, i.e. Group A: sentences containing elem-

ents in lexical contrast; Group B(i): sentences containing a new element realised as the leftmost lexical item; Group B(ii): sentences containing a new element realised as the rightmost lexical item and Group C: sentences containing elements of equal status according to the 'given' and 'new' structure. Let us examine each of these groups of sentences.

Group A: consists of the following sentences (each containing one element in lexical contrast, underlined):

Exp. 3. 3. ITS NOT THE YOUNGEST (ABC) SON WHO DOES IT
0 19
(3) (3)

Exp. 4. 3 7

Exp. 5. 17

Exp.3. 5. IS IT THE MILLER'S(ABC) DAUGHTER WHO IS
 15 0
 (7) (1)

Exp. 4. 10

Exp. 5. 17

	VERY	BEAUTIFUL
Exp. 3.	0	0
	(2)	(4)

Exp. 3. 23. THE $\frac{\text{DOG}}{22}$ (ABC) RUNS AWAY

Exp. 4. 10

Exp. 5. 17

All of the above sentences contain one element in lexical contrast and that element is identified as the tonic element by every judge. When a judge selects more than one element as tonic (which happens only in experiment 3) then the word which is in lexical contrast is selected as the 'primary' or most important tonic. The above experiments show that when the surrounding information structure is available to the judges the results are absolutely consistent.

An element which is in a situation of lexical contrast (i.e. which contradicts an element in the previous utterance) is realised on the cumulation of phonetic maxima (ABC) which is identified consistently as the tonic element. In sentence 3 in experiment 4 (the reading experiment) not enough context has been provided for the judges to identify youngest as the element in lexical contrast (the various sons have been discussed previously) therefore the judges select part of the whole constituent youngest son, but this is the only variation in tonic identification for these three sentences in experiments 4 and 5 which present the context surrounding an utterance containing a contrasted element.

Group B (i): consists of the following sentences, each of which contains a new element as leftmost lexical item (the new element is underlined)

Exp. 3.	15.	WAS (C)	IT	THE	<u>BAD (B)</u>	<u>NATURED (B)</u>	<u>DAUGHTER (ABC)</u>
					0	0	17
					(2)	(1)	(5)

Exp. 4. 9
(1)

Exp. 5. 12
(5)

Exp. 3.

	WHO RAN AWAY
	0
	(4)

Exp. 4.

0
(1)

Exp. 5. 0
(5)

Exp. 3. 17. THE $\frac{\text{HORSE (ABC)}}{15}$ REFUSES (BC) TO CROSS
(5) 0
(1)

Exp. 4.

	3	
	(3)	1

Exp. 5. 13
(4)

Exp. 3.	A	STREAM	AND	IS	SHOT	(B)
	0				2	
	(2)				(5)	

Exp. 4. 0 3
 (1) (3)

Exp. 5. 1
(4)

Exp. 3. 25. THEREFORE IT MUST BE THE SECOND (ABC) SON
0 0 7
(2) (1) (15)

Exp. 4.

	0	7
	(1)	(3)

Exp. 5. 5
(12)

Exp. 3.	WHO MARRIES THE MILLER'S DAUGHTER	0 (8)	0 (4)	0 (2)
Exp. 4.			0 (2)	
Exp. 5.		0 (3)	0 (11)	
Exp. 3.	29. WAS IT THE <u>SECOND</u> (ABC) SON WHO MARRIED	11 (10)	0 (1)	0 (2)
Exp. 4.		7 (3)		
Exp. 5.		10 (7)		
Exp. 3.	A MILLER'S DAUGHTER	0 (1)	1 (7)	
Exp. 4.		0 (2)	0 (1)	
Exp. 5.		0 (2)	0 (5)	

All of the above sentences contain a new element as leftmost lexical item, and these elements are identified as tonics by at least 19 out of 22 judges. When judges select more than one tonic, the new element is judged to be the most important or 'primary' tonic. The results for the above group of sentences are not as clear as the results for group A, reflecting the fact that there are 'extra' phonetic signals in the above sentences. In the sentences in group A, there was only one realisation of each phonetic maximum realised as the cumulation (ABC) on the contrasted element. In the sentences in group B (i), there are examples of several phonetic maxima within one sentence. Thus a new element realised as the leftmost lexical item is not identified as the tonic as consistently as a contrasted element.

Group B (ii): consists of the following sentences, each of which contains a new element realised as rightmost lexical item or constituent (the new element is underlined)

Exp. 3. 19. WAS THE OLD (A) MAN (BC) VERY WELL OFF

5	3	8
(6)	(3)	(5)

Exp. 4. 1 1 8

Exp. 5.	1	1	11
	(2)	(3)	(4)

Exp. 3. 27. SO IT WOULD BE THE RICH FARMER (ABC) WHO
0 0 7
(1) (5) (12)

Exp. 4.

	0	5
	(1)	(4)

Exp. 5.

	0	4
	(2)	(13)

Exp. 3. HAD THREE SONS

 0 0

 (4) (6)

Exp. 4. 0
(5)

Exp. 5. 0 0 0
 (1) (2) (10)

Exp. 3. 39. IS THE BAD (ABC) NATURED DAUGHTER VERY BEAUTIFUL
1 1 12
(7) (3) (7)

Exp. 4.	0	1	8
	(1)		(1)

Exp. 5.

0	1	2	9
(3)	(2)	(2)	(5)

The above group of sentences show interesting results. Sentences 19, and 36 are the most consistent sen-

tences of the group, demonstrating clearly that when the judges only have the written context available, they select an element from the new constituent in the utterance. When the judges are presented with the phonetic signal of the sentence without the surrounding context, the judgments are evenly spread out between the various stressed syllables in the sentence. But when the phonetic content of the utterance and the context is available to the judges, then the majority (11 out of 16 in both cases) select some element in the new constituent as the tonic despite the fact that the phonetic signals indicate the rightmost lexical constituent of the sentence as being more prominent. Thus judgments change according to the information available if there are several elements competing as tonic. In the above sentences, the phonetic maxima would predict that the theme should be the 'focus' or tonic of the utterance, whereas the information structure would predict that the rheme should contain the 'focus'. Predictably, then, the results indicate that when only the phonetic criteria are available to the judges, there is confusion, but when the context is presented as well, then the confusion is partially resolved and the judges tend to select a single tonic in the new constituent of the utterance.

Sentence 39 conforms to the above conclusions apart from one judge who selects bad-natured as the tonic in

the reading experiment. This choice might be supported since in fact it is the parent of the bad-natured daughter which is given in the preceding discourse, not the bad-natured daughter itself, thus the constituent bad-natured daughter may well be treated as 'new' according to the preceding discourse structure.

Sentence 27 is a little different from the other sentences in this group. Although rich farmer has been mentioned in the preceding discourse, the constituent is being introduced into a new relationship, i.e. with three sons. Since the relationship between the two constituents is new, perhaps both constituents have 'equal status' in some sense. This is certainly reflected in the results where the majority of judges select both elements as equal tonics, especially when the context is available. Thus although rich farmer might be analysed as 'given' in that it is recoverable from the preceding discourse, the speaker treats the constituent as 'new' since it occurs in a new relationship.

In summary, when a 'new' element occurs in an utterance, judges tend to select that element as the tonic of the utterance. If the new element occurs as the leftmost lexical constituent of the sentence, the new element is selected as tonic with or without context. If the new element occurs as the rightmost lexical constituent of the sentence, the context is necessary for the new element

to be identified as the tonic.

Group C: consists of the following sentences, each of which contains elements which are equal in status according to the preceding discourse structure.

Exp. 3. 2. THE DOG(ABC) RUNS AWAY(B)
 5 1
 (16) (16)

Exp. 4. 3 5
 (2) (2)

Exp. 5. 0 0
 (17) (17)

Exp. 3. 11. IS IT(AB) THE RICH FARMER(BC) WHO HAD
 0 0 1
 (1) (3) (14)

Exp. 4. 1
 (4)

Exp. 5 2
 (13)

Exp. 3. THREE SONS(B)
 0 3
 (3) (15)

Exp. 4. 1 4
 (4)

Exp. 5. 0 2
 (1) (12)

Exp. 3. 12. ALL(AC) THE SONS(BC) MENTIONED ARE SONS(BC)
 3 0 0 0 0
 (11) (1) (6) (2) (10)

Exp. 4. 1 0 0
 (2) (2) (1)

Exp. 5. 1 0 0 0
 (12) (1) (6) (5)

Exp. 3. OF THE SAME FATHER

1
(11)

Exp. 4. 0 4
(1) (4)

Exp. 5. 1
(12)

Exp. 3. 14. WAS THE OLD (AB) MAN (ABC) VERY WELL OFF
5 3 0 4
(5) (7) (1) (7)

Exp. 4. 1 2 0 4
(1) (2) (1) (2)

Exp. 5. 0 4 2
(2) (10) (11)

Exp. 3. 28. IS IT THE DOG (ABC) THAT RUNS AWAY (AB)
4 0 7
(11) (1) (11)

Exp. 4. 9 1

Exp. 5. 1 5
(11) (11)

Exp. 3. 30. THE YOUNGEST (ABC) SON DECIDED (B) TO
4 2 0
(10) (6) (8)

Exp. 4. 2 1
(1) (2)

Exp. 5. 1 0 1
(13) (3) (11)

Exp. 3. GET MARRIED
0
(8)

Exp. 4. 5
(1)

Exp. 5. 1
(3)

The above sentences all contain elements which are equal in status according to the preceding discourse structure. All of the sentences contain a spread of phonetic maxima realised on the various stressed items in each sentence, and each stressed item is selected on at least one occasion as a tonic. The results within each sentence correspond very closely across experiments, such that the majority of judges select the same items as tonic. For example in sentence 2, 17 out of 22 judges in experiment 3 decide that there are two equal tonics on dog and away, each realised on phonetic maxima; in experiment 4, where there are no phonetic cues, the judges choose one tonic selecting either of these two elements; in experiment 5, where the judges have access to both the phonetic information and the preceding discourse, all of the judges select both elements as tonic. The trend visible in experiments 3 and 4 therefore seems to culminate in experiment 5. But although a distinct trend is discernable in experiments 3 and 4, judges tend to make different kinds of decisions depending on the information available to them, and if there is a conflict between the kinds of decisions being made in the two experiments, the phonetic signals outweigh the contextual signals, thus reflecting the kinds of decisions the speaker has made about the utterance. For example in sentence 2 mentioned above, judges in experiment 3 tend to select two tonics, but in experiment

4, they tend only to select one tonic. In experiment 5 where all of the information is available, all of the judges select two tonics, reflecting the decision taken when only phonetic information was available.

There is perhaps more obvious conflict in sentence 28. The results of experiment 3 indicate that the majority of judges (12 out of 22) select two tonics, but that the majority of the remainder who select only one tonic select the rightmost lexical item. In experiment 4, 9 out of 10 judges select the leftmost lexical item which is the identifier of a cleft sentence with runs away given immediately previously. According to the preceding discourse therefore dog is selected as the tonic element. When all of the information is available to the judges in experiment 5, 11 out of 17 decide there are two equal tonics, and the majority of the remainder (5 out of 6) select the rightmost lexical item as the single tonic. The conflict is therefore resolved in favour of the phonetic signal.

Conclusions

Halliday's position, that tonic is the realisation of information focus, seems to be too strong. When the information focus is a contrasted element which contradicts a previous element, then the information focus is identified consistently as tonic. When the information focus is a 'new' element realised as the leftmost lexic-

al item then it is normally identified as tonic (i.e. by a majority). When the information focus is a 'new' element realised as the rightmost lexical item in a sentence more than one tonic is selected if the sentence is presented out of context (one tonic is selected on the given element which is in leftmost position, and one tonic is selected on the new element in rightmost position). If a new element in rightmost position is presented in context, then the new element will tend to be identified as tonic. Thus different decisions are taken when a sentence is presented out of context and when a sentence is presented in context. If judges are asked to identify tonics in utterances which do not have a clear 'given / new' structure, several tonics are identified.

These results create problems for Halliday's proposal that an unmarked tone group contains a single tonic which in its unmarked form will occur on the rightmost lexical item of a clause. Of the sentences listed under Group C, at least 5 out of the 6 would be regarded as one clause by Halliday (i.e. excluding sentence 12) since in 1967 Halliday refers to It was John that painted the wall as a single clause, yet all of these 6 sentences contain at least 2 tonics according to the judges (note: these tonics are not realised as 13 or 53 but as consecutive falls). Sentence structure does not affect tonic identification since all of the above groups, which dis-

play consistent results within the groupings, have contained a variety of sentence types including declaratives, cleft sentences and questions of various kinds. All 7 sentences which do contain only one tonic in examples 3, 4 and 5 (as identified by a majority of judges) have the leftmost lexical item as Halliday (and Crystal) would predict. I would like to propose that those utterances which are identified as containing a single tonic in the series of experiments are 'marked' in that the 'tonic' is a boosted peak of prominence -- boosted to add emphasis to that particular element (e.g. in a contrastive situation, or on a 'new' element). A normal, unmarked utterance will contain several peaks of prominence realised on the stressed elements of the utterance.

In order to preserve the notion of one 'tonic' per unit, I would like to propose that there is an abstract unit called a tone unit which contains only one stressed syllable and any preceding unstressed syllables (plus any following unstressed syllables of a polysyllabic word -- cf. chapter 8). The tone unit will normally be co-terminous with the foot (cf. Abercrombie, 1967). Indeed it may be unnecessary to differentiate between the two. Tone units may then combine to form tone groups. The stressed syllables of the tone group will interact according to the context of the utterance, the inform-

ation structure of the surrounding discourse, the information structure of the utterance itself and the speaker's intent. For example in a contrastive utterance the peak on the contrasted element will be boosted and the remaining peaks will be suppressed so that there may only be one 'perceived' peak of prominence (a tone group may therefore contain several perceived peaks of prominence or a single perceived peak of prominence). The tone group in its unmarked form will contain several peaks of prominence.

Appendix 1: results of experiment 1

Numbers of judges who select more than 1 tonic are bracketted.

S	ABC	AB	BC	AC	A	B	C	LLI	Other
1		pound 19 (7)					change 3 (7)		
2	sons 24 (3)			three 2 (3)					
3				change 5 (4)		pound 20 (4)			
4	there 12 (3)							house 14 (3)	
5		heard 10 (10)		pleased 9 (9)					were 0 (1)
6		there 2					my 0	house 27	
7					old 0 (3)	posted 7 (5)	man 0	letter 15 (4)	asked 2 (2)
8				played 1 (1)		squash 4 (6)		Fridays 6 (4)	M, W, F. 10 (3)
9	three 27							sons 2	
10					I 1	game 20 (1)	may 4 (1)		watch 3 (1)

Appendix 1: results of experiment 1.

S	ABC	AB	BC	AC	A	B	C	LLI	Other
11	many 1							have 1	children 27
12					lost 11 (4)	each 3 (2)	games 3 (9)	played 3 (4)	
13	played 9 (1)							M.W. & Fridays 9 (4)	squash 7 (5)
14		how 4 (2)	children 16 (4)			have 2 (1)		though 0 1	many 2 (1) you (1)
15			game 12 (3)		I 10 (3)				watch 4
16	play 27 (2)							me 0 (1)	do 0 (1)
17					how 0	children 27	many 0	have 1	you 1
18			daughters 18 (4)	three 7 (3)					I 0 (1)
19	pleased 27 (1)							it 1	heard 0 (1)
20	lost 4 (6)							played 0 (2)	each 1 (2) games 17 (6)

Appendix 2

Experiment 1: results.

1. HAVE YOU CHANGE (C) OF A POUND (AB)
 3 19
 (7) (7)
2. THREE (AC) SONS (ABC)
 2 24
 (3) (3)
3. HAVE YOU CHANGE (AC) OF A POUND (B)
 5 20
 (4) (4)
4. THERE (ABC) IS MY HOUSE
 12 14
 (3) (3)
5. HOW PLEASED (AC) THEY WERE WHEN THEY HEARD (AB) IT
 9 0 10
 (9) (1) (10)
6. THERE (AB) IS MY (C) HOUSE
 2 27
7. THE OLD (A) MAN (C) ASKED IF SHE HAD POSTED (B) THE
 0 2 7
 (3) (2) (5)
 LETTER
 15
 (4)
8. HE PLAYED (AC) SQUASH (B) ON MONDAYS, WEDNESDAYS AND
 1 4 16
 (1) (6) (7)
 FRIDAYS
9. THREE (ABC) SONS
 27 2
10. MAY (C) I (A) WATCH THE GAME (B)
 4 1 3 20
 (1) (1) (1)
11. HOW MANY (ABC) CHILDREN DO YOU HAVE
 1 27 1

12. HE LOST(A) EACH(B) OF THE GAMES(C) THEY PLAYED
 11 3 3 3
 (4) (2) (9) (4)

13. HE PLAYED(ABC) SQUASH ON MONDAYS; WEDNESDAYS AND
 9 7 9
 (1) (5) (4)
 FRIDAYS

14. HOW(AB) MANY CHILDREN(BC) HAVE(B) YOU THOUGH
 4 2 16 2 0 0
 (2) (1) (4) (1) (1) (1)

15. MAY I(A) WATCH THE GAME(BC)
 10 4 12
 (3) (3)

16. DO YOU WANT TO PLAY(ABC) WITH ME
 0 27 0
 (1) (2) (1)

17. HOW(A) MANY(C) CHILDREN(B) DO YOU HAVE
 27 1 1

18. I HAVE THREE(AC) DAUGHTERS(BC)
 0 7 18
 (1) (3) (4)

19. HOW PLEASED(ABC) THEY WERE WHEN THEY HEARD IT
 27 0 1
 (1) (1)

20. HE LOST(ABC) EACH OF THE GAMES THEY PLAYED
 4 1 17 0
 (6) (2) (6) (2)

Appendix 3.

Experiment 1: results.

2. THREE (AC) SONS (ABC)

2	24
(3)	(3)

9. THREE (ABC) SONS

27	2
----	---

16. DO YOU WANT TO PLAY (ABC) WITH ME

0	27	0
(1)	(2)	(1)

19. HOW PLEASED (ABC) THEY WERE WHEN THEY HEARD IT

27	0	1
(1)	(1)	

5. HOW PLEASED (AC) THEY WERE WHEN THEY HEARD (AB) IT

9	0	10
(9)	(1)	(10)

4. THERE (ABC) IS MY HOUSE

12	14
(3)	(3)

6. THERE (AB) IS MY (C) HOUSE

2	27
---	----

13. HE PLAYED (ABC) SQUASH ON MONDAYS, WEDNESDAYS AND

9	7	9
(1)	(5)	(4)

FRIDAYS

8. HE PLAYED (AC) SQUASH (B) ON MONDAYS, WEDNESDAYS

1	4	16
(1)	(6)	(7)

AND FRIDAYS

20. HE LOST (ABC) EACH OF THE GAMES THEY PLAYED

4	1	17	0
(6)	(2)	(6)	(2)

12. HE LOST (A) EACH (B) OF THE GAMES (C) THEY PLAYED

11	3	3	3
(4)	(2)	(9)	(4)

7. THE OLD (A) MAN (C) ASKED IF SHE HAD POSTED (B)
 0 2 7
 (3) (2) (5)

THE LETTER
 15
 (4)

1. HAVE YOU CHANGE (C) OF A POUND (AB)
 3 19
 (7) (7)

3. HAVE YOU CHANGE (AC) OF A POUND (B)
 5 20
 (4) (4)

10. MAY (C) I (A) WATCH THE GAME (B)
 4 1 3 20
 (1) (1) (1)

14. HOW (AB) MANY CHILDREN (BC) HAVE (B) YOU THOUGH
 4 2 16 2 0 0
 (2) (1) (4) (1) (1) (1)

17. HOW (A) MANY (C) CHILDREN (B) DO YOU HAVE
 27 1 1

11. HOW MANY (ABC) CHILDREN DO YOU HAVE
 1 27 1

15. MAY I (A) WATCH THE GAME (BC)
 10 4 12
 (3) (3)

18. I HAVE THREE (AC) DAUGHTERS (BC)
 0 7 18
 (1) (3) (4)

Appendix 4.

Join in the search for tonics!

The TONIC of a sentence has been defined by many different people in many different ways. Could you please state below the criteria you yourself would use to define the TONIC of a sentence.

There is then a short questionnaire to find out details of your experience in Phonetics and your accent background. Would you please complete the questionnaire.

You will find the experiment itself on the attached sheet.

- A. The criteria I would use to define the TONIC of a sentence would be
-
-
- B. 1. NAME
2. NATIONALITY 3. NATIVE LANGUAGE
-
4. FLUENCY IN ANY OTHER LANGUAGE/S
5. KNOWLEDGE OF ANY OTHER LANGUAGE/S
6. a) NATIONALITY OF MOTHER
6. b) NATIVE LANGUAGE OF MOTHER
7. a) NATIONALITY OF FATHER
7. b) NATIVE LANGUAGE OF FATHER
- C. Please state qualification and / or experience in Phonetics

Appendix 5.

Below is a short story. Read it through and try to remember it. Your partner will ask you questions about the events and characters in the story, but you should only answer 'yes' or 'no'. While you are being asked questions, you may refer back to this typed copy.

HOW A BAD DAUGHTER WAS MADE A GOOD WIFE

Once there was a rich farmer who had three sons.

The eldest of them married a gentleman's daughter, who was very well off; and the second married a miller's daughter who was very beautiful. And the youngest of them, who was a carpenter, was then the only one left unmarried. He decided to marry.

'I'll get the daughter of a bad mother, as bad as I can find.'

He took his pony and his dog, and went away, not knowing how far he would go. The day was bad, and snow was falling. What did he see on his way but an old man at work ploughing in a field. He went over to him, and said:

'Oh, ho! you've got a bad day for ploughing!'

'Well,' said the old fellow, 'indeed, it isn't good!'

'Why do you have to work outside on a day like this?'

'If there was a way I could stay at home, I wouldn't come out myself; but I'd rather be outside, than indoors with the womenfolk. If her mother is bad, my daughter is seven times worse.'

'May I marry her?' asked the young man.

'I never saw anyone on whom I'd wish her, but if you think you can bring her to heel, I'll not keep her from you.'

'Oh, I'll take her right enough, if you'll give her to me.'

'That's just what I'll do,' said the old man.

So they agreed to get married; and the daughter of the old man became the wife of the farmer's youngest son. Then they left for his home. The pony was their only conveyance, and the girl was put behind him on the pony's back.

They left, and they hadn't gone far before the young husband's dog ran away. He shouted at the dog, but the dog paid no heed. When the dog did come back, he took a revolver out of his pocket and shot the dog and killed it.

'Goodness,' said his wife, 'why did you kill the dog?'

'Why should I let him live, a worthless creature that wouldn't listen to me?'

When they were getting near to his home, they came to a river which ran between them and the house. He put the pony at the worst bit of the stream, and began to lash the pony to make it go across the river. The pony only backed and would not go near it. He asked his wife to dismount, and he got off the pony himself, and took out

his revolver, and put it to the pony, and killed it.

'My God, why did you kill the pony?' she said.

'Why should I let it live, a worthless creature that wouldn't do what I told it?'

He went and took the saddle off the pony.

'Here,' he said. 'You carry that.'

'Indeed I'll not carry it; you might have let the beast that was carrying it live; the saddle would be fitter on it than on me.'

'Are you saying you won't carry it?' he said.

'Oh, of course I'll carry it, but you really might have let the pony live, it would have been more suitable on her.'

'Well, if the pony had done what I told her, I wouldn't have killed her.'

They came home, he and she, and she was the best wife there had ever been! There was nothing he asked her to do that she wouldn't do!

Appendix 6.

This is a game of cooperation. Your partner has been given a story to read and remember. Below is a list of characters from the story (List A) and a list of actions or events from the story (List B). Your task is to match up List A and List B and reconstruct the story.

Your partner has already read the story. You may ask him / her questions about the story, but your partner may only answer 'yes' or 'no'. So, you will not be able to ask questions like: "Who had three sons?" You may ask as many questions as you like until you think you know which characters performed which actions and in what order, that is, until you think you can tell the story. (Note: One character may perform more than one of the actions in List B).

When you feel you are able, tell the whole story to your partner.

If you wish, you can make notes, draw connecting lines, or number actions on this sheet of paper.

List A

(Characters)

1. eldest son
2. youngest son
3. old man
4. miller's daughter
5. bad-natured daughter

List B

(Actions/Attributes)

- a) _____ has a bad-natured wife and a daughter who is worse.
- b) _____ does not obey command to return and is shot
- c) _____ had three sons

List A

(Characters)

- 6. dog
- 7. pony
- 8. newly-married couple
- 9. rich farmer
- 10. second son
- 11. gentleman's daughter

List B

(Actions/Attributes)

- d) _____ is impressed by previous events and becomes a very obedient wife.
- e) _____ returns home on the back of a pony.
- f) _____ asks to marry bad-natured daughter.
- g) _____ was very beautiful.
- h) _____ went off with his pony and dog to look for the bad-natured daughter of a bad-natured mother
- i) _____ refuses to cross a stream and is shot.
- j) _____ was very well off.
- k) _____ runs away.
- l) _____ married a gentleman's daughter.
- m) _____ decided to get married.
- n) _____ tells his wife to carry the saddle.
- o) _____ meets _____ ploughing a field in the snow.
- p) _____ married a miller's daughter.

Appendix 7.

Experiment 2: results.

1. THE PONY(ABC) RUNS AWAY
25
2. WAS IT THE MILLER'S DAUGHTER(ABC) WHO RETURNS HOME
9 0
(7) (3)
ON THE BACK OF THE PONY
9
(7)
3. PRESUMABLY IT WAS THE ELDEST(ABC) SON
25
4. THE SECOND(ABC) SON MARREID THE MILLER'S DAUGHTER(AB)
2 12
(11) (11)
5. WAS IT THE YOUNGEST SON(ABC) WHO TELLS HIS WIFE TO
8
(11)
CARRY THE SADDLE
6
(11)
6. DID THE RICH FARMER(ABC) HAVE THREE SONS
23 0
(2) (2)
7. WAS IT THE SECOND(AC) SON WHO MARRIED THE MILLER'S
12
(7)
DAUGHTER(B)
6
(7)
8. THE DOG(ABC) RUNS AWAY
23 2
9. WAS IT ONE OF THE SONS(ABC) WHO WAS VERY WELL OFF
23 0
(2) (2)
10. WAS IT THE RICH FARMER'S ELDEST SON(ABC) WHO MARRIED
23
(2)

THE GENTLEMAN'S DAUGHTER

0
(2)

11. IT WAS THE ELDEST(ABC) SON

25

12. WAS IT THE YOUNGEST(ABC) SON THAT MARRIED THE VERY

20

(5)

0

(1)

BEAUTIFUL MILLER'S DAUGHTER

0
(4)

13. WAS IT THE MILLER'S(ABC) DAUGHTER WHO WAS VERY

23

(2)

BEAUTIFUL THEN

0
(2)

14. WAS IT THE WIFE(AC) OF THE NEWLY-MARRIED COUPLE(B)

6

(12)

5

(13)

WHO WAS IMPRESSED(B) BY PREVIOUS EVENTS(A) AND

0

(3)

0

(4)

BECAME A VERY OBEDIENT WIFE

0
(1)

15. WAS IT THE BAD-NATURED DAUGHTER(ABC) WHO RAN AWAY

21

(2)

2

(2)

16. IS IT THE MILLER'S(ABC) DOG

25

17. WELL DID THE OLD MAN(ABC) HAVE THREE SONS(AB)

0

(2)

5

(11)

7

(11)

18. IS IT THE PONY(ABC) WHO REFUSES TO CROSS A STREAM

11

(8)

0

(3)

AND IS SHOT(AB)

5
(8)

19. WAS IT THE RICH FARMER (Bc) WHO HAD THREE SONS (A#)
10 8
(7) (7)

20. SO THEY GOT MARRIED
1 24

Appendix 8: results of experiment 2

S	REF	ABC	AB	BC	AC	A	LLI	Other
1	CONTRAST	pony 25						
3	CONTRAST	eldest 25						
6	CONTRAST	farmer 23 (2)					sons 0 (2)	
8	CONTRAST	dog 23					runs away 2	
11	CONTRAST	eldest 25						
16	CONTRAST	millers 25						
20	CONTRAST	got 24						so 1
10	CONTRASTED CLEFT	rich farmer's eldest son 23 (2)					daughter 0 (2)	
13	CONTRASTED CLEFT	millers 23 (2)					beautiful 0 (2)	
15	CONTRASTED CLEFT	bad-natured daughter 2 (2)					away 2 (2)	

Appendix 8: results of experiment 2

S	REF	ABC	AB	BC	AC	A	LLI	Other
2	CLEFT	miller's daughter 9 (4)					pony 9 (4)	home 0 (3)
5	CLEFT	youngest son 8 (11)					saddle 6 (11)	
9	CLEFT	sons 23 (2)					well off 0 (2)	
12	CLEFT	youngest 20 (5)				married 0 (1)	daughter 0 (4)	
18	CLEFT	pony 11 (8)	shot 5 (8)					stream 0 (3)
7	CLEFT				second son 12 (4)		daughter 6 (4)	
14	CLEFT		impressed 0 (3)		wife 6 (12)		obedient 0 (1)	couple 5 (13) events 0 (4)
19	CLEFT			rich farmer 10 (4)		sons 8 (4)		
4	NON-CON NON CLEFT	second 2 (11)	miller's daughter 12 (11)					
17	NON CON NON CLEFT	old man 5 (11)	three sons 7 (11)					did 0 (2)

Appendix 9.

(Experiments 4 and 5)

You will find below, 17 excerpts from a 'game' situation. Each excerpt contains one sentence which has been underlined.

You are asked to decide which word in the underlined sentence you would classify as the tonic of the sentence. Please encircle the word you choose.

(1) 1 No the eldest son marries a gentleman's daughter.

Youngest son marries the bad-natured one therefore
it must be the um the second son who marries the
milller's daughter

2 Yes

1 But the question is whose second son

(2) 1 Is it the gentleman's daughter who was very beautiful

2 Um No

1 Is it the miller's daughter who is very beautiful

2 Yes

(3) 1 Well then at this point we're going to have to have
the dog does not obey the command to return and is
shot. Sounds totally implausible

2 Well he runs away first

- 1 The dog runs away. Is that the next thing he does?
- 2 Yes
- 1 Extraordinary story - twelve - and then he does not obey the return and is shot
- 2 Yes
- 1 It does seem harsh
- 2 Well yes
- 1 Which really frightens the girl. He then at this point tells his wife to carry the saddle
- 2 No
- 1 The horse refuses to cross a stream and is shot
- 2 Yes
- 1 I see
- (4) 1 The father or the parent of the bad-natured daughter is presumably very well well off
- 2 Is that a question
- 1 Yes. Is the the parent of the bad-natured daughter very well off
- 2 No
- 1 No. Is the bad-natured daughter very beautiful
- 2 No. At least we're not told that
- 1 Well then it has to be a gentleman's daughter who is very beautiful
- (5) 1 Who meets somebody ploughing a field in the snow. Is it the newly-married couple who meet somebody ploughing a field in the snow
- 2 No

- 1 No. Its not the youngest son who does it
- 2 Yes
- 1 The youngest son who does it
- (6) 1 Is it the rich farmer who has a bad-natured wife
and a daughter who is worse
- 2 No
- 1 Is it the old man who had a bad-natured wife and
a daughter who is worse
- 2 Yes
- 1 (Whispers .. had three sons .. bad-natured wife...)
Is it the rich farmer who had three sons
- 2 Yes
- (7) 1 Did the old man have a bad-natured wife and a
daughter who is worse
- 2 No. Oh yes. Yes he did
- 1 Right. So there's an old man who has a bad-natured
wife and a daughter who is worse. Um. Does the old
man have sons as well
- 2 No
- 1 He doesn't
- (8) 1 The newly-married couple return home on the back of
a pony. The dog runs away. It doesn't obey the
command to return and is shot.

(9) 1. Somebody ran away. Was it the bad-natured daughter who ran away

2 No

1 No

(10) 1 Did the miller have a bad-natured wife and a daughter who was worse

2 No

1 Did the old man have a bad-natured wife and a daughter who is worse

2 Yes

1 Was the old man very well off

2 No

1 Was the miller very well off

2 No

(11) 1 Well. You tell me the next thing she said firmly

2 Well. Right. The next thing he does is is that he not only I mean he goes off to look for a bad-natured daughter. He for some reason decides to marry a bad-natured daughter.

1 Oh I see. Well that's very curious

2 Yes

1 So O.K. So that's going to be the youngest son decided to get married

2 uh huh

1 and seven then went off with his pony and dog to look for the bad-natured daughter of a bad-natur-

1 ed mother

2 uh huh

1 It does seem a curious desire to have in life

(12) 1 Refuses to cross a stream and is shot. Its gonna have to be that. How extraordinary. How about the miller's eh the bad-natured daughter refuses to cross a stream and is shot It can't possibly be

2 No

1 They're going to have to do something like go to market and I can't find anything saying going to market. How about the horse runs away

2 No

1 The dog runs away

2 Yes

1 Is than ten eleven

2 No

1 But it is going to be the dog who runs away sooner or later. Well that's something isn't it.

(13) 1 But was it the eldest son who decided to get married

2 No

1 Was it the second son who decided to get married

2 No

1 Then it was the youngest son who decided to get married

2 Yes

1 Married a miller's daughter. Now who was it who

1 had three sons. The rich farmer had three sons.

Was it the second son who married a miller's daughter

2 Yes

(14) 1 Is the person who has a bad-natured wife and daughter who is worse the rich farmer

2 No

1 Is it the old man

2 Yes

1 So it would be the rich farmer who had three sons

2 Yes

(15) 1 Is that the do- pony which does not obey the command to return and is shot

2 No

1 Is it the dog that does not obey the command to return that is shot

2 Yes

1 O.K. Runs away. Then it's it must be is it the dog that runs away

2 Yes

1 How shall I ever unravel this set of railway lines. That's the dog

(16) 1 Does the yongest son go off with his pony and dog to look for the bad-natured daughter of a bad-natur-

1 ed mother

2 Yes

1 Eh where are we. Eh youngest son two Um Was the old man very well off

2 No

1 Was the rich well was the rich farmer very well off

2 Um No

(17) 1 Now who on earth married the miller's son. That was the other daugh- uh son that's the um eldest son isn't it. Uh No the eldest son marries a gentleman's daughter. The youngest son marries the bad-natured one therefore it must be the um the second son who marries the miller's daughter.

2 Yes

1 But the question is whose second son. Um. The second son. Now the Oh I see all the sons mentioned are sons of the same father

2 Yes

1 Right. I get it. Yes.

Chapter 6.

6.0 An initial examination of contour units in readings of a text

In chapter 5 it was clearly demonstrated that judges did not choose consistently between competing cues for tonic identification. If there was a confusion of phonetic cues such that two elements X and Y seemed likely candidates as 'tonic', judges would be divided and select either X as a single tonic, or Y as a single tonic, or both X and Y as two tonics within the utterance. Judges did not consistently select the rightmost 'moving' deviation from the baseline as the 'nuclear' movement as Halliday and Crystal might predict. Throughout chapters 4 and 5, it has been noted that there are several deviations from the baseline within a pause-defined unit in Edinburgh English, and that these deviations are often transcribed as being equal in height. Chapter 5 attempted to find out whether one of these deviations would be more likely to be perceived as prominent, but this was not the case. If a confusion of phonetic cues existed, then choice existed, and judges selected any or all of the phonetically 'prominent' items as tonic (i.e. as the most prominent item).

In this chapter I propose to examine pause-defined units in a text with specific reference to the acoustic analysis of the fundamental frequency characteristics of such units (see chapters 1 and 2 where F_0 is said to be the closest acoustic correlate of intonation). These

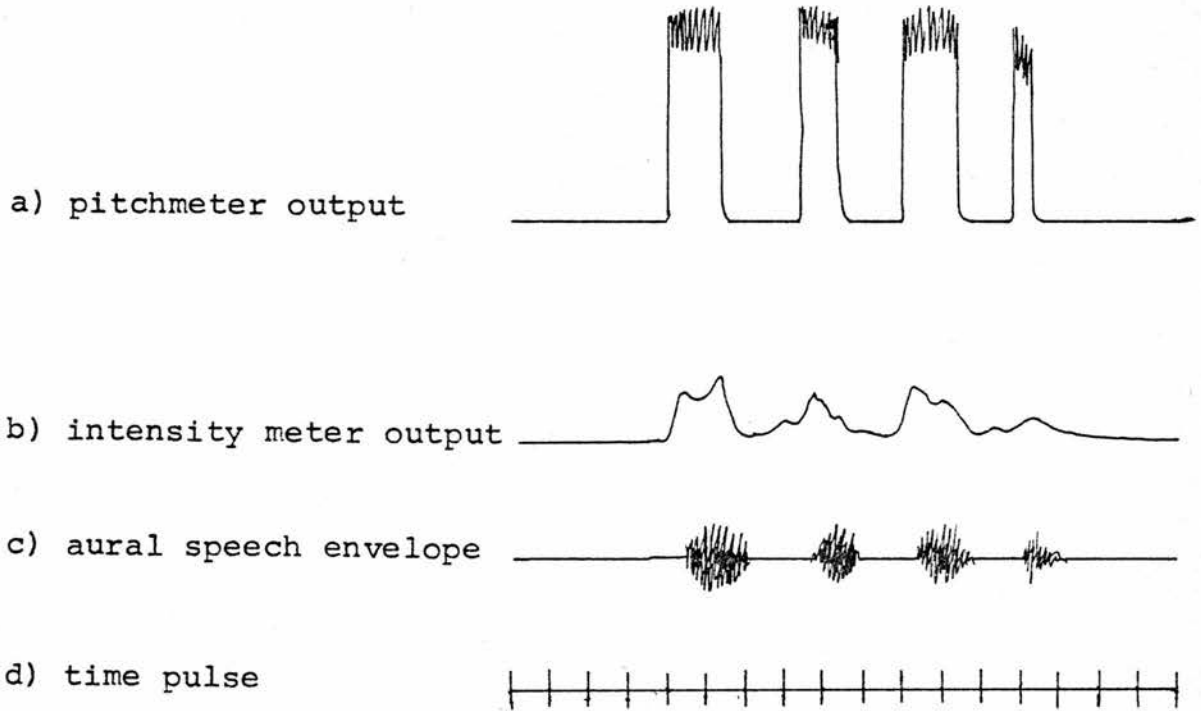
units were examined initially within the readings of a text since this allowed a comparison to be made of the same lexical content spoken by different speakers. An examination of spontaneous speech would not have allowed this, and would have involved false starts, hesitation pauses, incomplete sentences, partial repetitions, filled pauses and other common performance variables (see chapter 4 for a more detailed account of these variables).

Six speakers were asked to read the text 'The Witch's Daughter' (see 4.1.1). All six speakers were native speakers of Scots and all were born and brought up in Edinburgh. Three speakers were male, three were female; one male and one female were in the 20-30 age-group; one female was in the 30-40 age-group; one male and one female were in the 40-50 age-group; and one male was in the over 50 age-group. The occupations of the speakers were 1 student, 1 librarian, 1 laboratory technician, 1 housewife, 1 works engineer, and 1 retired gentleman.

The six readings were fed through the Frøkjaer-Jensen Transpitchmeter and the Frøkjaer-Jensen Intensity Meter simultaneously, and the results were combined on the output of a 4-track Mingograph. On one sheet of Mingograph paper, therefore, the four lines reading from top to bottom represented a) the pitchmeter output; b) the intensity meter output; c) the overall speech envelope

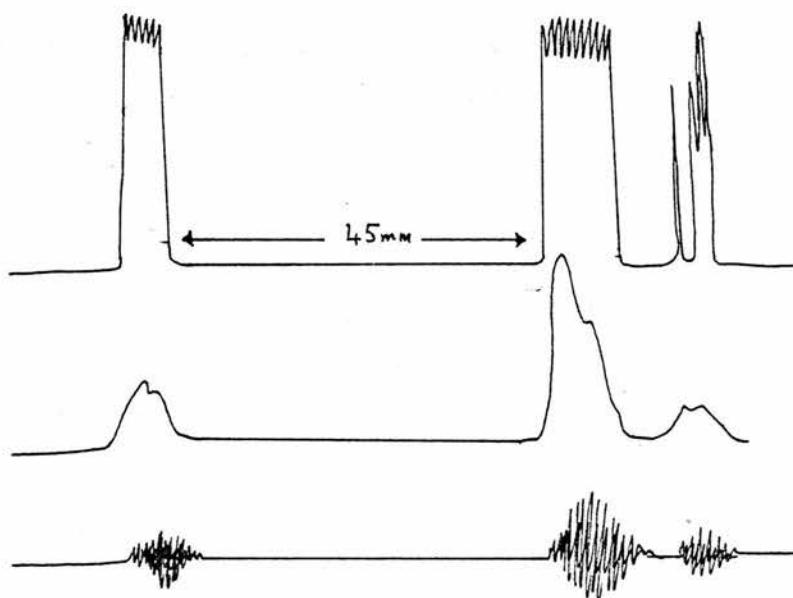
(i.e. the 'unadulterated' speech signal); and d) a time pulse (see Fig. 1):

Fig. 1



Graphs showing the pause duration between units were drawn up comparing the six speakers. Pause measurement was carried out from the Mingograph output where the pitchmeter reading, the intensity reading and the voice signal all returned to their respective baselines indicating silence, cf. Fig. 2:

Fig. 2 (showing a pause of 0.9 secs)



This output was then converted to a numerical reading as follows in Fig. 3, with pause measurement given in mm (millemetres) where 1 cm=0.2 sec. The pause duration is placed vertically under the position where it occurs in the text:

Fig. 3

an aged widow

in white clothes

20

17

12

20

28

20

18

24

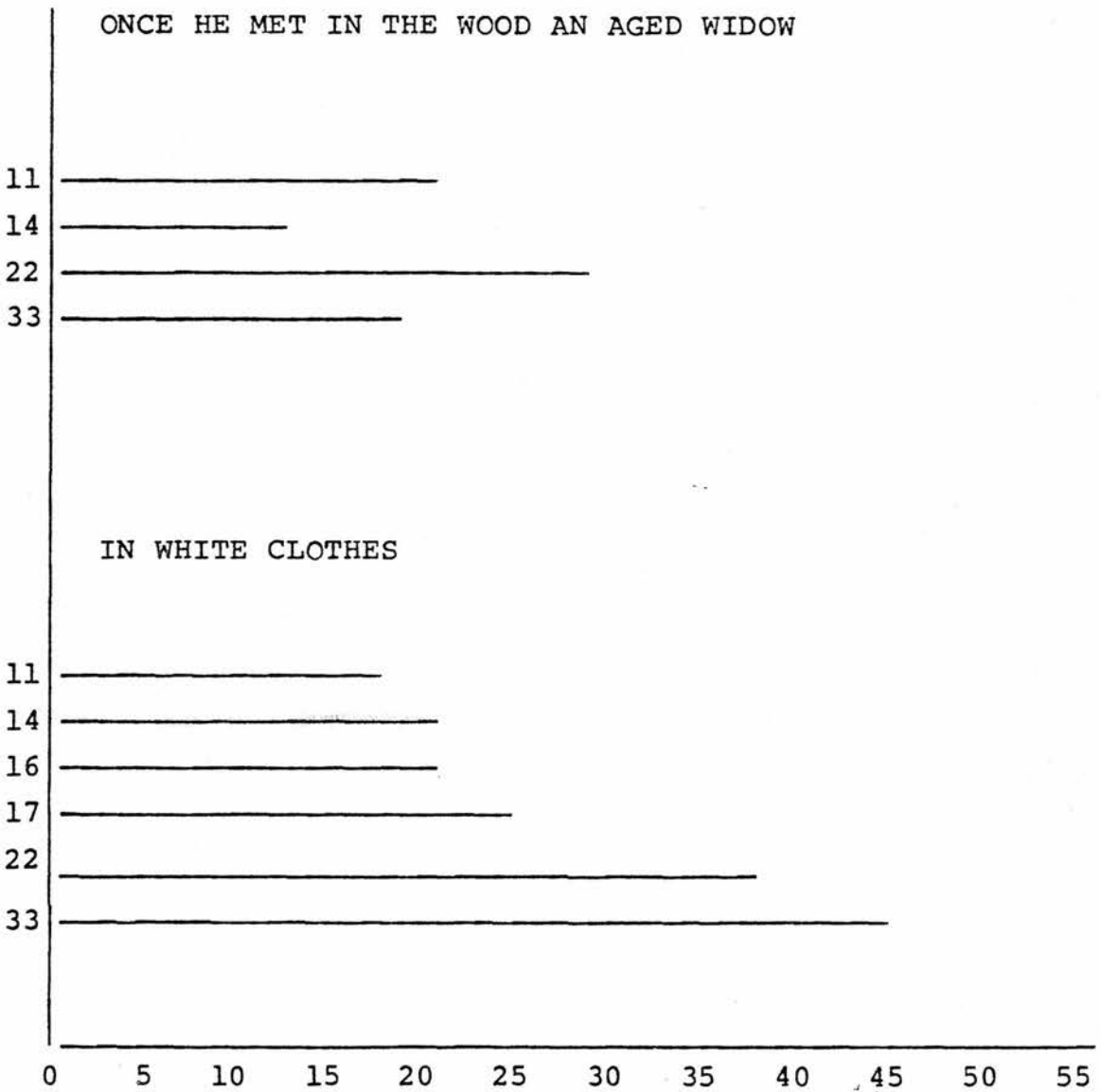
37

44

These figures were then converted into the following kinds of graphs (the pause duration is inserted at the point where the pause occurs in the text):

Fig. 4

subj.
number

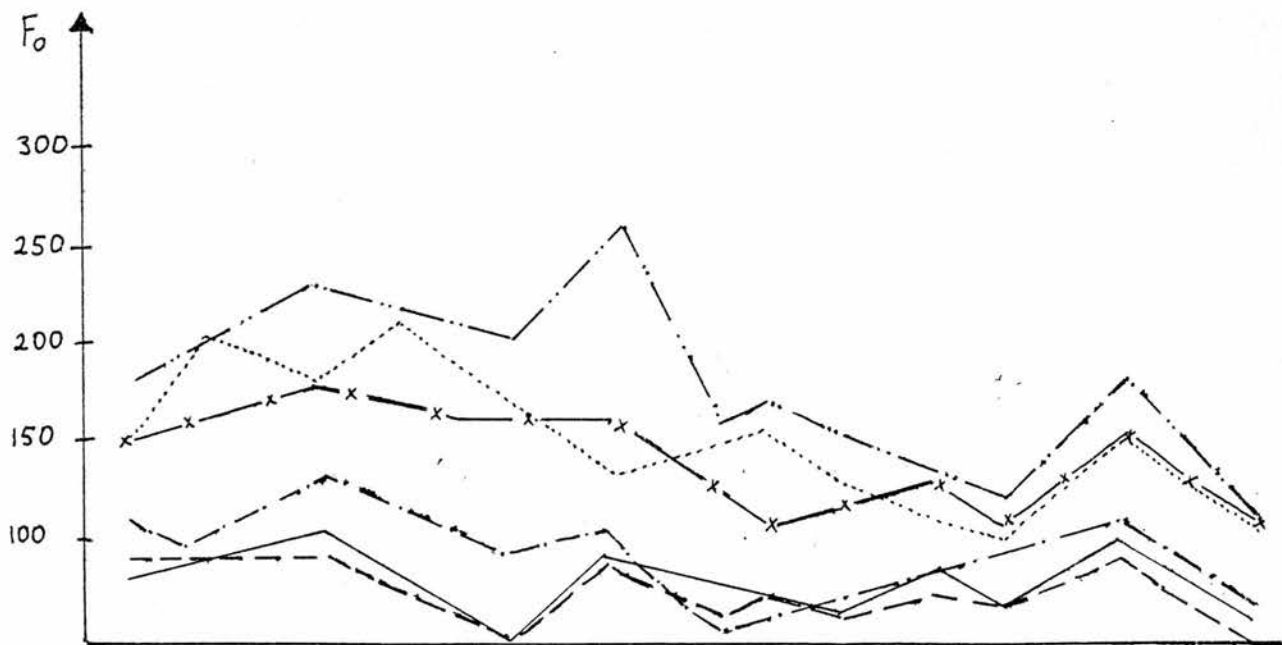


I then measured the Fundamental Frequency (F_0) of each pitch movement in cycles per second (cps) and drew graphs representing the pitch movement from these figures. From the characteristics of the F_0 wave form, the intensity wave form and the wave form of the total voice signal, the F_0 measurements were assigned to the words of the text. The assignment of these figures was also supported by auditory analysis. Thus Fig. 5 became Fig 6:

Fig. 5

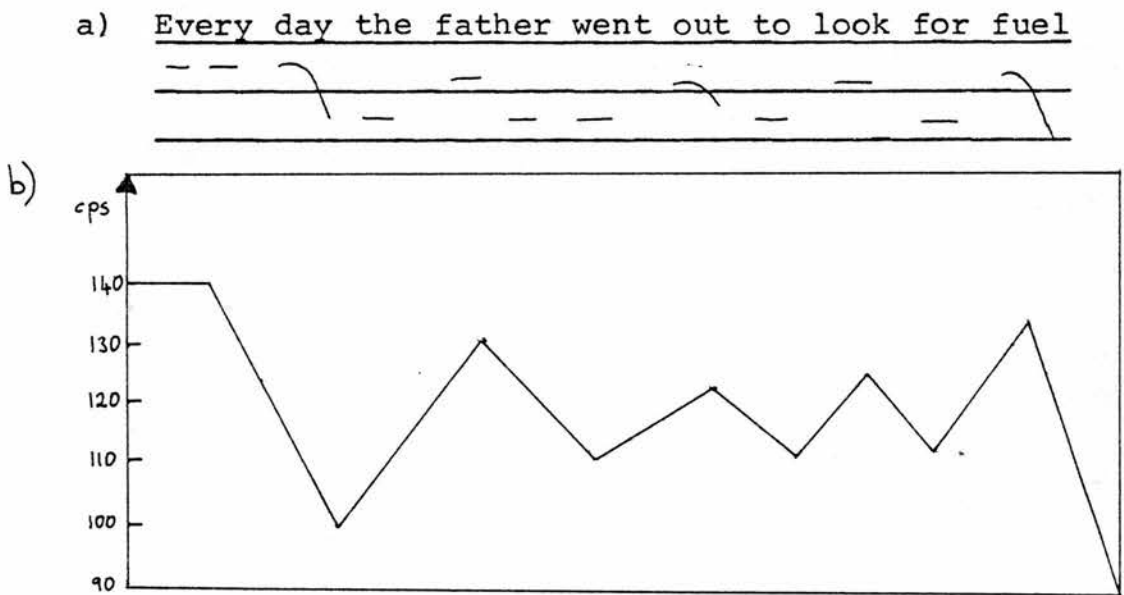
Every day the father went out to look for fuel					
200-225-210	210	160	180-160	200-160	
200-250-225	260-225	225-180	200-170	160	200-160
230-275-250	300-200	210-190	170	200	245-160
140	140-100	130-110	120	110-120-115	130- 95
130-150-100	140-110	110	135-115	145-110	
160-150	175-135	140	120	110-140	160-120

Fig. 6



Conventions: in Fig. 5, figures which are joined together by a hyphen represent continuous movement, for example 200-225-210 represents a rise-fall occurring over the phrase every day for the first speaker. The extremes were measured as the perceptual primes, because the auditory analysis indicated that the extremes were the main parameters used for auditory transcription i.e. for auditory perception. Thus if we compare one of the graphs drawn from the acoustic data with its auditory equivalent, we find that the two match very closely, see Fig. 7:

Fig. 7



In Fig. 6, words which are in capitals contain a peak of prominence i.e. deviate from the baseline. Words

which are underlined contain peaks for at least 5 out of the 6 speakers. Words which are bracketted appear to be optional peaks and contain peaks for less than 5 out of 6 speakers. The pause duration graphs were then used in order to determine, in part, which 'units' were to be examined as 'units' in the Fo graphs.

6.1 Pause duration measurements

All of the pause measurements included in these graphs are measurements of boundary pauses i.e. pauses define units with semantic or grammatical cohesion (normally both). A full set of graphs is given in Appendix 1. Different lengths of pause recurred frequently indicating a fairly regular patterning of pauses throughout the text.

The longest pause occurs for the majority of the speakers (i.e. 4 out of 6) after 'The old man stroked his beard for a while but finally gave his assent'. The remaining two speakers produce their longest pause before this sentence. For four out of the six speakers this longest pause measures approximately 1.24secs, whereas for the other two, who pause for longer periods more often throughout the text, the pause is approximately 1.9secs long. There is also an equivalent 'long' pause preceding this utterance measuring approximately 1.22 secs. Therefore this utterance seems to be being used as some kind of thematic 'bridging' device operating between two main themes. The utterance occurs approximate-

ly halfway through the text and separates the lead up to the game of chess from the results of the game. Pause duration, therefore, seems to signal that some kind of thematic 'paragraphing' is taking place. Lehiste (1975) notes that paragraphs possess a suprasegmental structure which indicates the beginning and end of paragraphs and characterises the body of the paragraph. She specifically examines the fundamental frequency correlates of this hypothesis but notes that 'similar relationships are expected to prevail with regard to timing and stress' (p. 195). Such 'cognitive' paragraphs need not only occur at points of orthographic paragraphing. Indeed if pauses of a certain length (i.e. over 1.0sec) can be said to indicate places where 'paragraphing' has taken place, such pauses occur regularly in the six readings of the text in the following places (where the utterance preceding the pause is given, and the pause is indicated by ++):

- a) ...where an old man lived with his three sons ++
- b) ...every day the father went out to look for fuel ++
- c) ...who was seated on a square stone playing chess ++
- d) ...'Certainly', said the old man ++
- e) ...'How many children have you though' ++
- f) ...you must send me your sons as sons-in-law ++
- g) ...but finally gave his assent ++

h)...and again after three days the youngest ++

i)...how pleased they were when they heard it ++

At least 3 out of 6 speakers choose to paragraph at all of these points, only three of which co-incide with an orthographic paragraph (i.e. at points (b), (g) and (i)).

These pauses will be referred to as long pauses since they consist of the longest pauses which recur regularly throughout the readings of the text. The first three long pauses occur after each of the first three sentences of the text. These sentences are fairly long and each contains new information which is not available to the speaker. Each sentence could therefore be regarded as introducing new thematic material which will constitute 'topics' throughout the text. The first three pauses occur in the text as follows:

a) In the midst of a range of wild mountains, was a

small straw hut, where an old man lived with

his three sons ++

b) Every day the father went out to look for fuel ++

c) Once he met in the wood an aged widow in white clothes who was seated on a square stone playing chess ++

The next long pause occurs after the first convers-

ation between the old man and the woman i.e. after the first section of direct speech. This pause therefore occurs in the following position in the text:

- d) Since the old man was a keen player himself, he asked the woman, "May I watch the game?" She replied, "Do you want to play with me?" "Certainly," said the old man. ++

There is then a sentence in indirect speech which introduces the next thematic section. The fifth long pause occurs part way through the next conversation between the old man and the aged widow preceding the discussion of the bet. This pause occurs as follows:

- e) When the widow asked for what stakes they should play, he suggested playing for his wood. But the old woman said, "No, we can't play for wood because I don't have any wood. How many children have you though?" ++

The sixth long pause occurs after the discussion of the bet has been completed, thus occurring after the following section of text:

- f) "Three sons", was his answer. "Three sons? That is perfect. I have three daughters. If you win

I will send them as brides for your three sons;
but if I win, you must send me your sons as
sons-in-law." ++

The next long pause has been discussed previously. It is the longest pause in the text for 4 of the 6 speakers, and seems to function as some kind of turning point in the text dividing the settling of the bet from the outcome of the game. This pause then, occurs after the following sentence:

g) The old man stroked his beard for a while, but finally gave his assent. ++

The eighth long pause occurs after the instructions have been given for the bet to be fulfilled, as follows:

h) He lost each of the games he played, and when the widow got up to leave she said, pointing down into a dark valley, "There is my house. Tomorrow send me your eldest son, three days later the second, and again after three days the youngest." ++

The next long pause co-incides with an orthographic paragraph and precedes the last sentence of the text

which might be said to operate as a summary of the result of the whole episode described in the text. This pause occurs after the days events have been communicated to the sons of the old man, as follows:

- i) She then departed, and the old man went home without collecting any more wood to tell his sons what had happened. How pleased they were when they heard it! ++

There is then a pause at the end of the text following the last sentence:

- j) The next day he sent the eldest son; three days later the second; and on the sixth day he sent the youngest.

All of these long pauses, then, occur after a specific theme or 'topic' is finished, dividing the text into clear thematic sections as demonstrated above. The thematic sections vary in length which indicates that the long pause is not merely a linear 'chunking' device which might occur for example after every four tone units. Each thematic section can be seen to be a 'cognitive' whole in some sense.

Long pauses occur regularly throughout the text and can be said to signal that cognitive paragraphing is

taking place. The long pauses had a duration of approximately 1.0-2.0 secs., and constituted the longest pauses produced consistently by the informants at tone unit boundaries. Pauses of a shorter duration also occurred consistently. These pauses measured between 0.4secs and 0.8secs and demarcated smaller units of speech than the long pauses. The units demarcated by these shorter pauses were then examined on the fundamental frequency graphs (Fo graphs).

6.2 Fundamental frequency characteristics

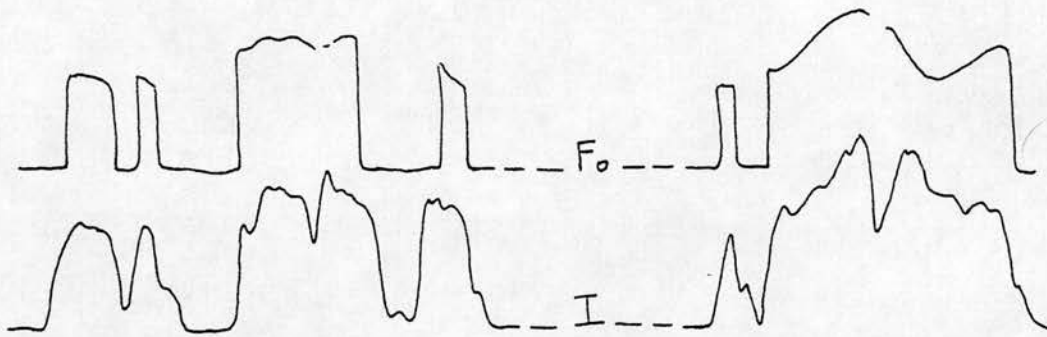
Units demarcated by long pauses or by the shorter pauses which we have referred to as boundary pauses will be referred to as contour units, since each unit contained a 'contour' i.e. a constantly varying fundamental frequency parameter. Lieberman states in 1965:

"Intonation is ... perceived in terms of complete contours of fundamental frequency and amplitude, i.e. ensembles of fundamental frequency functions and amplitude variations as functions of time."

This section will only deal with the characteristics of the fundamental frequency, however, since preliminary measurements of the amplitude output indicated that the fundamental frequency measurements and the amplitude

measurements were very closely matched, i.e. whenever there was a peak on the F_0 graphs, there was an equivalent peak on the amplitude graphs. If we examine a sample of the mingograph output which contains both the pitch-meter output (measuring the fundamental frequency) and the intensity meter output (measuring the amplitude) then the correlation is clearly seen:

Fig. 8



The contour units correlated in length with Crystal's tone units, or Halliday's tone group, all being approximately co-terminous with the clause. Crystal believes that the boundaries of tone units are demarcated by pauses (see chapter 3), thus there is ample support in the literature for examining the F_0 characteristics of pause-defined units.

The F_0 graphs can be examined in various ways:

- a) each graph can be examined individually,

- b) graphs can be made consisting of several phrases read by the same speaker,
- c) graphs can be made consisting of the same phrase as uttered by all six speakers.

(a) If the graphs are examined individually various patterns emerge which are illustrated in figs. 9 and 10. Figure 9 illustrates the general tendency for speakers to peak at the same height and to drop back to the same level. Figure 10 illustrates a second trend which is for speakers to fall a similar amount, but from progressively lower heights. These two trends are very common in the *Fo* graphs, but it is difficult to arrive at generalisations when examining a single unit produced by a single speaker.

(b) If graphs are superimposed upon each other such that several units spoken by the same speaker are compiled on the same graph, then graphs such as those in Figures 11 and 12 are obtained. Figures 11 and 12 are taken from the same speaker. Figure 11 consists of those units which occur as the initial syntactic clause in a sentence, preceding at least one other clause. Figure 12 consists of those units which occurred in the texts as final clauses following at least one other clause. There is no substantial difference between the contour shapes of these two groups of clauses apart from the fact that the

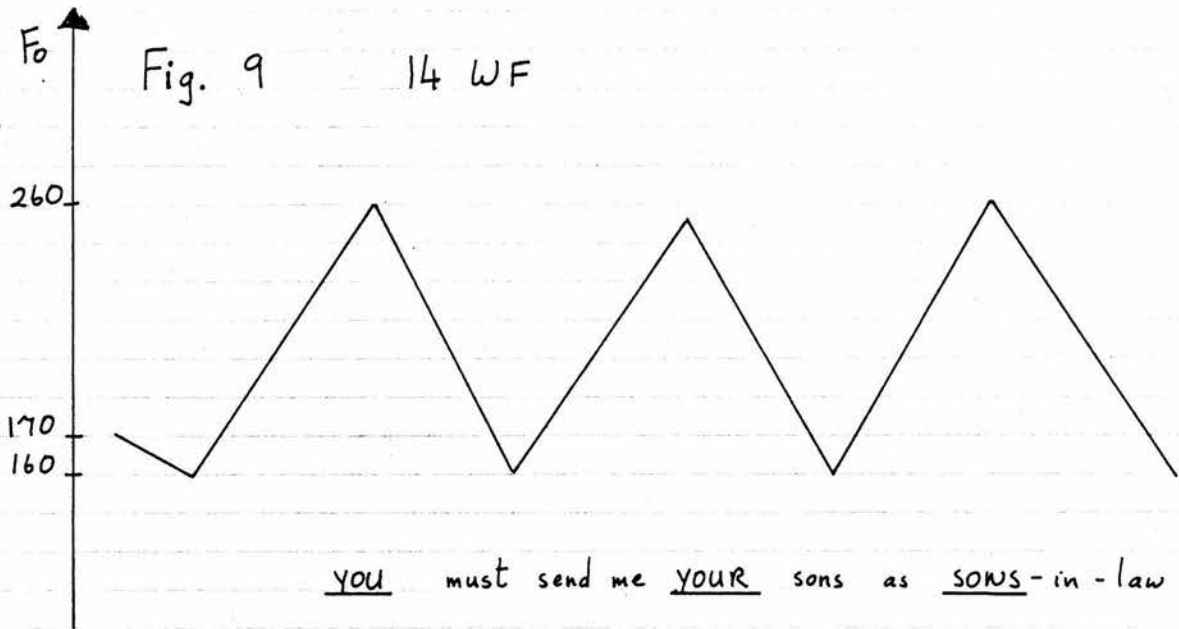
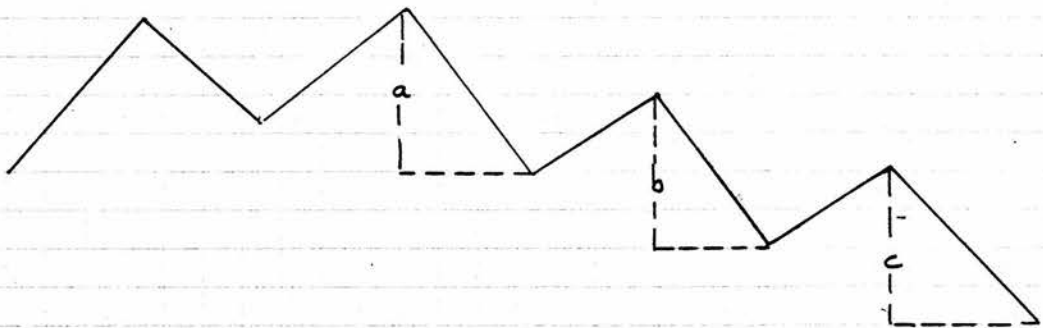


Fig. 10 33 WM



when the WIDOW ASKED for what STAKES they should PLAY,

where $a = b = c = 25 \text{ cps}$

(peaked elements are underlined)

Fig. 11 22BS INITIAL CLAUSES

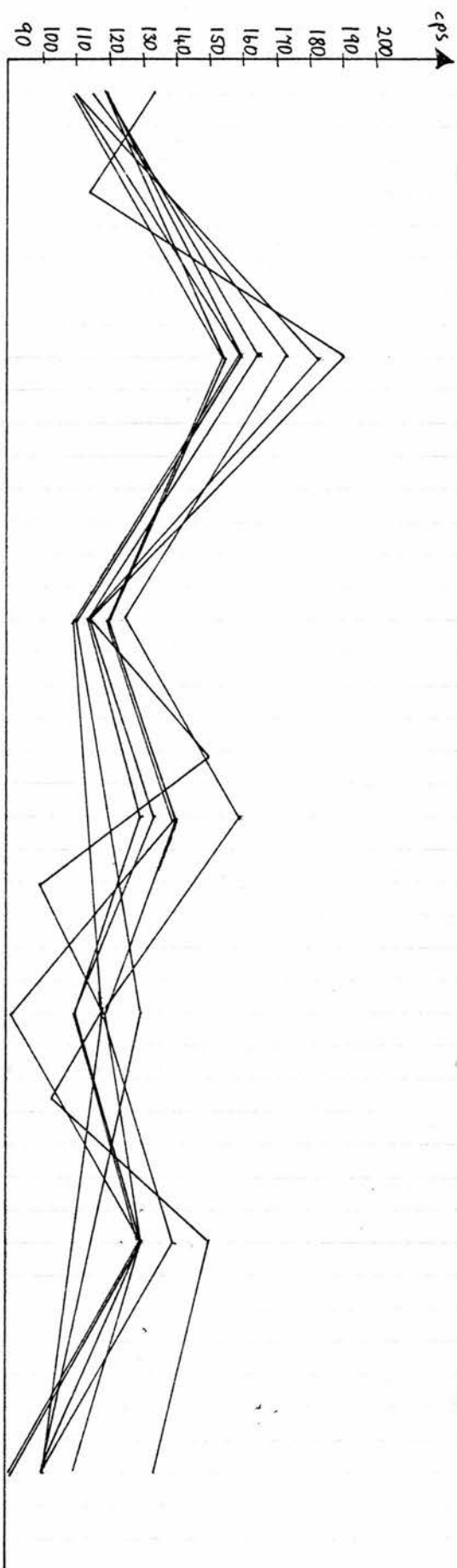
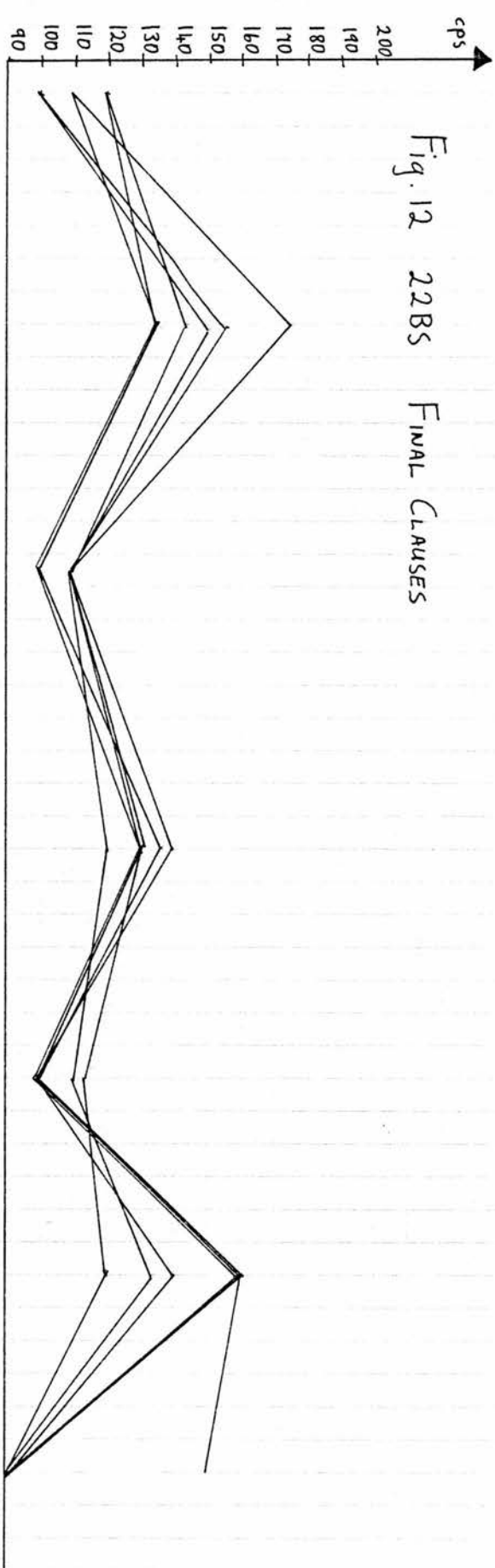


Fig. 12 22BS FINAL CLAUSES



initial clauses tend to consist of four peaks, whereas the final clauses tend to consist of three peaks. Apart from that, the only other difference between the two groupings of contour units is that the pitch range for the initial clauses is slightly higher than the pitch range for the final clauses, there being about 10cps difference between them. This difference can be illustrated as follows where the numbers represent the frequency ranges of the speaker's extremes:

Initial clauses

110-120 160-170 100-120 120-140 90-110

Final clauses

100-110 140-160 100-110 120-130 90-100

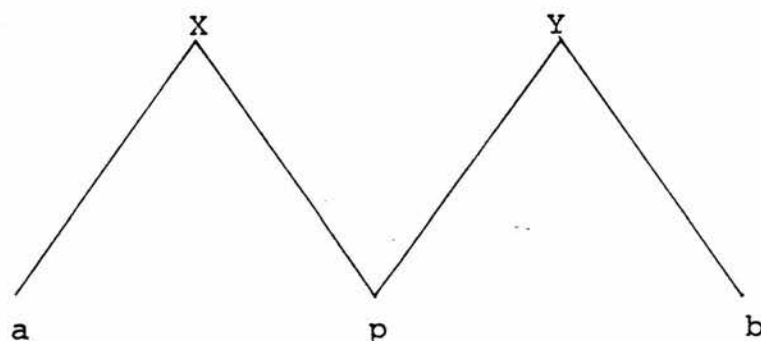
The contours of the final clauses therefore seem to be marginally lower in pitch range when compared with the initial clauses, but the shape of the contours is substantially the same showing a strong patterning of alternating high 'aiming' points, and low 'aiming' points. These aiming points will be referred to as peaks when they are high deviations from a 'baseline', and pivot points when they are low and seem to form part of a 'baseline'.

(c) The texts were then examined in detail accord-

ing to the Fo graphs which contained the same orthographic unit as spoken by all six speakers.

There were 266 contours altogether, roughly 44 units per text. Out of these 266 contours, 226 of them contained at least two peaks of prominence, 144 of these being double-peaked, and 82 having three or four peaks. The remaining 40 contours were single peaked contours containing only one peak of prominence. Fig. 13 is a diagrammatic representation of a double-peaked contour:

Fig. 13



When examining the fundamental frequency characteristics of the contour units, I will first examine the properties of the double-peaked contours since they are in the majority of the texts. I will then examine the three and four peaked contours to see if they can be related to the double-peaked contours.

Fig. 13 has been labelled as follows: a is the starting point of the contour; X is the maximum value of the first peak of the contour; p is a pivot point; Y is the maximum value of the second peak of the contour; and b is

the end point of the contour. (Any deviation from the baseline will be referred to as a peak).

The most obvious feature shared by the contours is of course the general shape of the contour, but there are also specific similarities common to the majority of the contours. The three points, a, p, and b, seem to form a very consistent baseline within each contour unit with a narrow range of variation of approximately 20-30 cps. at each point of the baseline.

If we examine the text of the speaker 11MR to find support for this notion of "baseline" within each unit, we find that out of a total of 44 units, all of which have beginning points (a) and end points (b), 27 out of 44 beginning points are between the frequencies 150-180 cps.; and 36 out of 44 end points are between the frequencies 140-160 cps. There are 31 contour units which have the pivot point (p) and all 31 pivot points are between the frequencies 150-160 cps. These figures would seem to indicate that the majority of the points a, p, and b occur within a pitch range of 140-180 cps. for the speaker 11MR. This proposed baseline a - p - b occurring within each contour unit might therefore be designated the 'normal' baseline since the majority of the points, a, p and b (94 out of a total of 119) occur within the limits mentioned above. The normal baseline for speaker 11MR can be said to have a pitch range of 40 cps. ranging between 140 and 180 cps. The pitch

range for the normal starting point being 150-180 cps.; the normal end point occurring within the range 140-160 cps. and the pivot point occurring consistently between the values 150-160 cps..

There are a few deviations from this 'norm' which has been set up and these deviations will now be examined in the light of the discourse.

The starting point can deviate from the norm by as much as 100 cps. when there is a new 'cognitive' paragraph (see Urquhart and Rees, 1976, Lehiste, 1975, and section 6.1 where long pauses are said to indicate paragraphing), e.g. speaker 11MR has a normal base at approximately 160 cps., but her starting point for the sentence The old man stroked his beard for a while is 270 cps. This admittedly is unusually high, and her normal frequency for a "key change" will be between 200 and 220 cps. (a key change being a noticeable change in pitch occurring at the beginning of a new 'cognitive' paragraph). The starting point, therefore, can be said to deviate from the norm at the beginning of a thematic or cognitive paragraph.

The end point (i.e. the Fo measurement preceding a pause) will deviate from the norm when there is either some continuity involved or when there is direct speech involved. Thus the end points will be higher in a 'continuation' contour e.g. in a clause which precedes a fin-

al clause e.g. was a small straw hut (where hut ends on 220 cps) preceding where an old man lived with his three sons (ending on 140 cps) and if you win (ending on 185 cps) preceding I will send them as brides for your three sons (ending on 160 cps), or in direct speech where the pitch range may vary according to the emotive content of the discourse.

The pivot points deviate from the norm least of the three baseline points, and the deviation tends to occur in the section of direct speech in the text and very occasionally in a continuation contour, otherwise the pivot points are the most regular of the three baseline points. An example of deviation on the pivot point is found in the text of speaker 16LK whose normal baseline pitch range is 170-200 cps., but in the contour How many children have you though? the pivot point is 245 cps. This unit occurs in direct speech and is a question. The whole contour unit is very high in the pitch range of speaker 16LK, and it is not therefore surprising that the pivot point is also high. This is the only deviation from the normal range by the pivot point in the text read by speaker 16LK and there are no deviations in the text read by speaker 11MR.

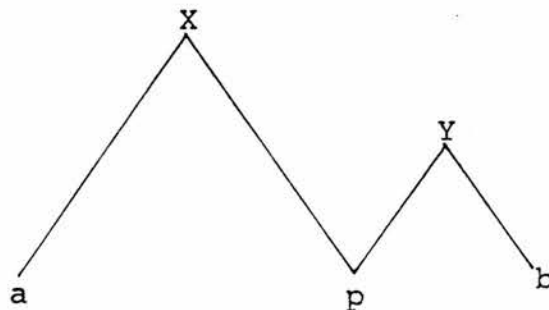
The 'normal' baseline of speaker 11MR therefore has a range as follows:

starting point	pivot point	end point
150-180 cps	150-160 cps	140-160 cps

with occasional deviations from this norm which can be accounted for in terms of the discourse. We can see from this that the starting point has the widest range, that the pivot point has the most narrow range and that the end point has the lowest range. These in fact are the features which distinguish these three points. The starting point has the greatest variety of frequency measurements, the end point is typically lower by 10-20 cps. than the starting point, and the pivot point is typically constant.

Having established a baseline a-p-b, let us now examine the peak points which deviate from the baseline, i.e. the points X and Y, and the relationship between them. There seems to be three typical patterns involved here. One where X is greater than Y; one where X is less than Y and one where they are deemed to be equal. Fig. 14 is a diagrammatic representation of the contour where X is greater than Y:

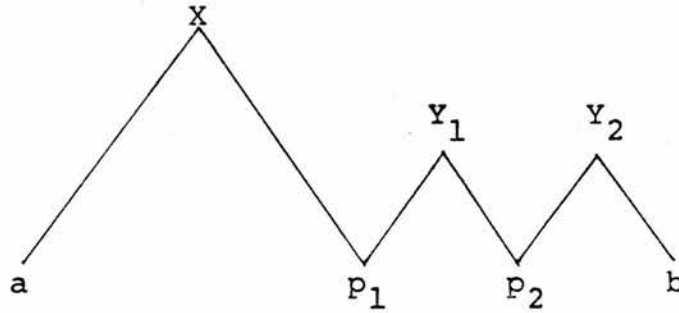
Fig. 14



This contour occurs regularly in the same kinds of places in all six texts, cf. Appendix 2. The texts have this contour for the units In the midst of a range of wild mountains, Since the old man was a keen player himself, The next day he sent the eldest son, the old man stroked his beard for a while, Every day the father went out to look for fuel, Once he met in the wood an aged widow, He lost each of the games he played, Tomorrow send me the eldest son, How pleased they were when they heard it, and on the sixth day he sent the youngest. All of these units except the last one are initial clauses, and may be said to be the initial sentence or clause of a cognitive paragraph or 'thematic section' each of which follows a long pause. Thus this kind of contour would seem to realise a key change in the discourse and will normally coincide with a higher starting point than the norm. Note Lehiste (1975) who found that 'High fundamental frequency appears associated with judgments that the sentence was produced at the beginning of the paragraph' (p. 198). Since the phonological or suprasegmental 'marking' of cognitive paragraphs seems to be fairly widely accepted (see Brazil 1978, Rees and Urquhart 1976, and Lehiste 1975) I will refer to these phonological units as paratones. The paratone will typically be surrounded by long pauses and the initial stressed syllable of a paratone will typically be high in the speakers' pitch range.

The final unit and on the sixth day he sent the youngest, was given the same contour as the initial contour of the story and may be functioning as a sort of 'rounding-off' contour but more investigation and comparison with other contours in final position is necessary before such a statement can be made with any force. The six speakers use this contour at various other places in the text not necessarily coinciding with each other depending upon the number of 'cognitive' units the speaker has divided the passage into. The contour also seems to coincide occasionally with adverbial phrases and clauses of time such as three days later the second, and when the widow got up to leave she said, so that the final unit of the story may be merely another example of a time constituent receiving the maximum peak in a contour unit. This, however, does not happen consistently, and more evidence is needed to support either or both of these arguments for the final unit of the story.

Let us now examine those units which have the maximum amount of movement on the initial fall as in Fig. 14, but which have more than two peaks. An example of this contour is given in Fig. 15, where X is greater than Y_1 or Y_2 .

Fig. 15

There is a marked similarity between those information units which have the contour 14, and those units which have the contour 15, in fact many of the information units are the same, i.e. one person selecting 14, another person selecting contour 15 for the same information unit. This high degree of agreement allows us to propose that contour 15 is also realised on the first unit of a new paratone. The proposal may even be extended to suggest that 14 and 15 are alternative contours, either of which may be realised at the beginning of a new paratone.

If we combine the information units which have either contour 14 or 15, as their realisations, we find that all six speakers have either contour 14 or 15 for

the units In the midst of a range of wild mountains, and Once he met in the wood an aged widow, we find that 5 out of 6 speakers have either contours 14 or 15 for the units Since the old man was a keen player himself, and when the widow got up to leave; and that 4 out of 6 speakers have contour 14 or 15 for the units Every day the father went out to look for fuel, The old man stroked his beard for a while, and How pleased they were when they heard it.

It is obvious then, that there is a high proportion of agreement between the speakers as to where the contours 14 and 15 should appear in the text. It would seem logical therefore to propose that these two contours are variations of the same contour, contour X, which is defined as a contour having the initial peak greater than all subsequent peaks within the contour.

Let us now examine the contour where X is less than Y, i.e. a contour of the following shape -- see Fig. 16:

Fig. 16

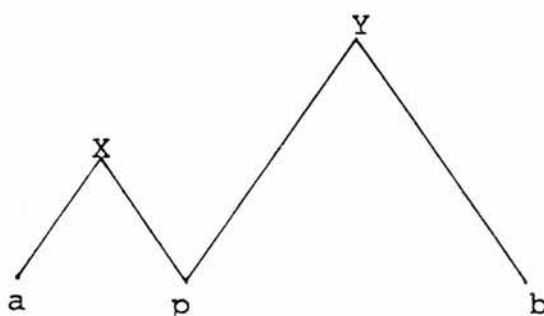
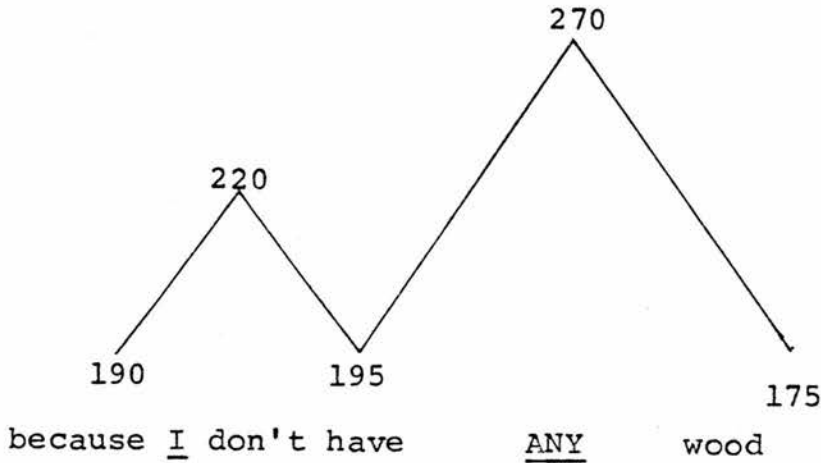


Fig. 16 (contd)

This contour is an example of the contour where X is less than Y. Speaker 14 WF (peaked elements are underlined)



This form seems to function in several ways and is realised in several different situations. This contour occurs mainly on information units containing a contrastive element towards the end of the unit, e.g. on units such as because I don't have ANY wood, I have three DAUGHTERS, three days later the SECOND, etc. (where the contrasted word is given in capital letters). Most of the contrastives occur during the section of direct speech although some occur in indirect speech. This contour also occurs on units which are final clauses such as the following: where an old man lived with his three sons, who was seated on a square stone playing chess, to

tell his sons what had happened and on direct questions such as Do you want to play with me and How many children have you though. Therefore, the contour with a final peak of prominence is realised in the text in the following situations:

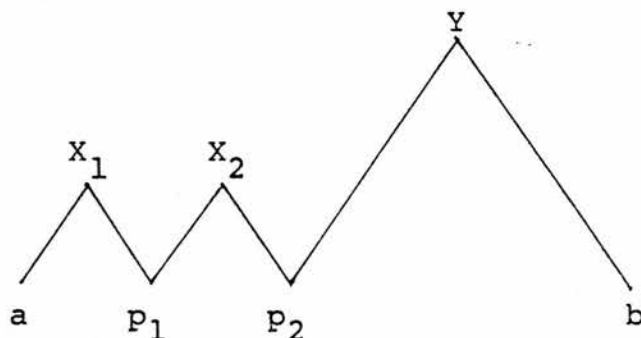
- a) when the information unit contains a contrastive element towards the end of the clause
- b) when the unit is a direct question
- c) when the unit is a final clause

It must be remembered, however, that this is a description of the situation where this contour arises, NOT a prescription predicting where this contour WILL arise. In fact the converse of the statement does not hold since this contour does NOT occur in all cases where there is a final clause, nor in all cases where there is a question. If we treat contrastives as 'different', in that a contrasted element can occur anywhere in an utterance and require a 'boosted' peak (as demonstrated clearly in chapter 5), then we might hypothesise about the similarity between the other realisations of this type of contour. If we accept that a question functions as a marker of 'end-of-turn', thus inviting another speaker to initiate his turn and reply to the question, the contour which is realised with a final

peak of prominence which is greater than the preceding one can be seen to function as a marker of 'finality' in a general way. Thus whereas the contour with an initial greater peak functions as a marker of 'initial' in a new paratone, the contour with a final greater peak can be said to function as a marker of 'final' in some sense.

Let us now examine the contour as exemplified in Fig. 17, where there is a final peak of prominence which is preceded by more than one smaller peak of prominence:

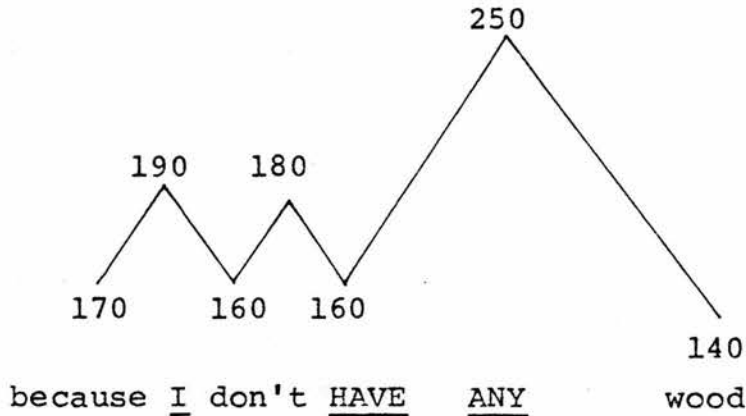
Fig. 17



There are only eleven examples of this contour in the six readings of the text, and 9 of these have a contrastive element present towards the end of the information unit as for example in Fig. 18, where the figures

given represent frequency in cycles per second:

Fig. 18



In this example ANY is the contrastive element and is realised by a final peak which is greater than the two preceding peaks by approximately 65 cps. Nine out of the eleven examples of this contour are similar to the one given in Fig. 18, i.e. having a contrastive element present towards the end of the information unit, the remaining 2 are final clauses. Therefore, just as contour 14 and contour 15 could be generalised as realisations of the same contour because they were realised on the same information units and occurred in the same kinds of places in the text, for the same reasons we can subsume contours 16 and 17 under the same heading as realisations of the same contour, contour Y, which is defined as having a final peak of prominence greater than any preceding peak of prominence.

We have now discussed contours where X is greater than Y and contours where X is less than Y, the next contour to be discussed is the contour where X seems to be in some way equal to Y. I say 'seems to be in some way equal' because although at least one parameter can be measured as having 2 equal values this does not necessarily mean that these values will be perceived as being equal (cf. chapter 2).

But since chapters 1 and 2 demonstrated that numerous experiments in the literature showed F_0 to be the strongest acoustic correlate of intonation as well as the most widely used correlate experimentally, this chapter will assume that equal F_0 measurements are, in some sense, perceptually equal. The intensity measurements throughout these texts peaked in the same places as the F_0 measurements thus supporting the F_0 measurements. Also, it was demonstrated in chapter 5 that although pitch movement is said to be a strong cue in perceiving nuclear movement, maximum pitch movement was not always selected as the nuclear movement. Thus we shall assume that the second hypothesis proposed in chapter 4 is correct, that peaks or deviations which are auditorily equal are perceived as equal and that pitch movement will be determined primarily by phonetic environment.

There were 226 contours from the 6 texts which had

more than one point of prominence, of these 226, 64 were deemed to have equal points of prominence according to the measurement of the fundamental frequency. This means that over one quarter of the contour units which had several points of prominence had equal points of prominence. Previous theories (cf. Crystal, 1969 and Halliday, 1963 1970) have consistently maintained that there is one and only one point of prominence within each tone group, the above measurements however indicate a very high proportion of contours having not only 2 point of prominence within one contour but 2 or more equal points of prominence within one contour unit. Admittedly few of these peaks of prominence are exactly equal, i.e. equal in every way, the majority of them are 'deemed' to be equal since the differences between the measurements of the fundamental frequency are very small -- averaging between 10 and 20 cps. Further experimentation must be carried out in order to ascertain whether 2 peaks of prominence which are found to be equal in F_0 are regularly perceived to be equal.

Let us now examine the contours within the 6 readings of the text which have equal peaks of prominence according to F_0 measurement. These units constitute approximately one quarter of the total number of units under discussion in this chapter, and may be subdivided into various headings according to the different ways in

which the peaks of prominence are related.

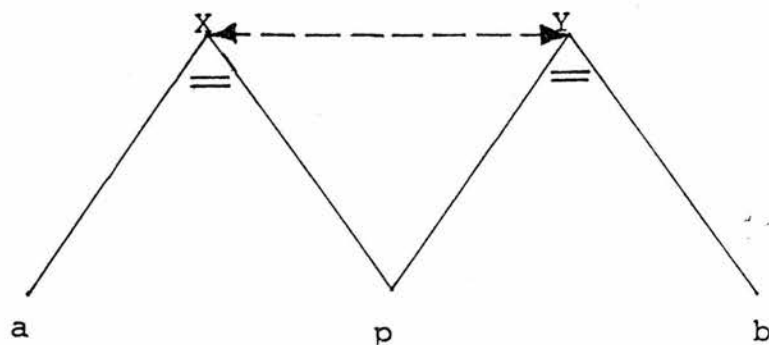
A) Absolutely equal

The contours under this heading contained two peaks of prominence which reached exactly the same height and fell to exactly the same fundamental frequency as each other, and can therefore be said to be absolutely equal. Examples are as follows: (where underlined elements contain the peak points):

	a	X	p	y	b
16LK <u>without</u> collecting <u>any</u> more wood	170	180	160	180	160
16LK <u>three</u> <u>sons</u>		250	210	250	210
11MR pointing <u>down</u> into a dark <u>valley</u>		165	160	165	160
11MR <u>you</u> must send me your <u>sons</u>					
as sons-in-law	160-265-160	260	160	260	160

and the contour would be of the shape in Fig. 19:

Fig. 19



It should be noted that the units are of different lengths (ranging from 2 lexical items to 8), and that the amount of pitch variation ranges from a rise or fall of 5 cps to a rise or fall of 100 cps.

B) Equal peak points

The contours under this heading had peaks of prominence which reached the same height but were followed by different amounts of fall -- the last fall normally falling more than the initial fall -- but not necessarily in every case. Examples are as follows:

	a	X	p	Y	b
17AC and <u>again</u> after <u>three</u> days	105	120	105	120	95
17AC he suggested <u>playing</u> for his <u>wood</u>	120	130	105	130	100
11MR how <u>pleased</u> they were when they <u>heard</u> it		180	225	170	225 160
33WM he <u>suggested</u> playing for his <u>wood</u>	120	175	120	175	140

There are two main types of contours under this heading cf. Fig. 20, but X was always equal to Y:

Fig. 20

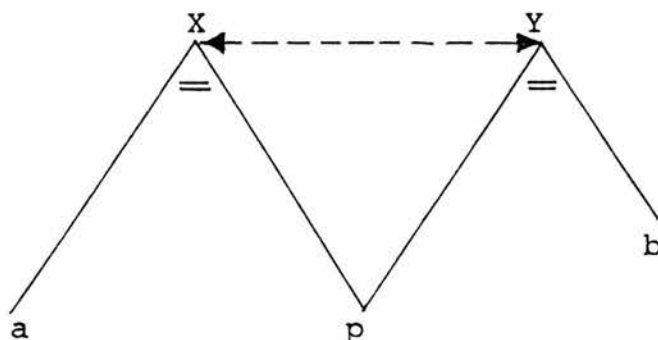
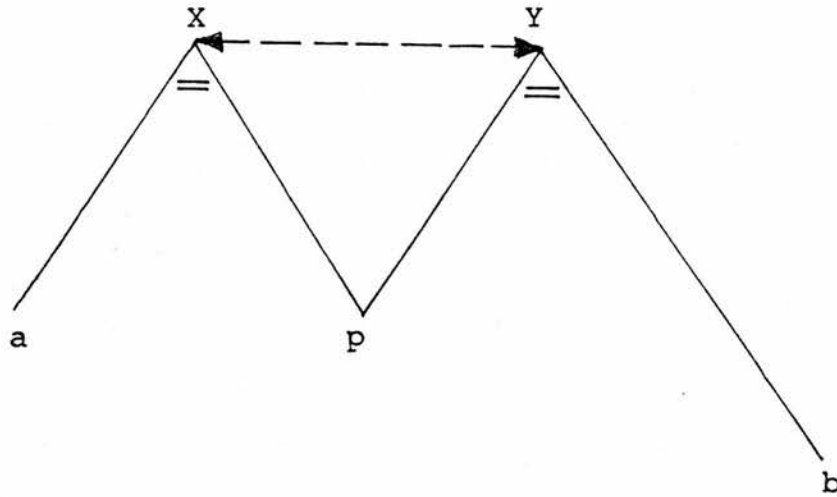


Fig. 20 (contd)C) Equal amount of fall in downward progression

The contours under this heading had peaks with an equivalent amount of fall but the final peak, Y, fell less than the initial peak, X. In other words the fundamental frequency of X was greater than Y, but the amount of fall from both was equal. The following are examples:

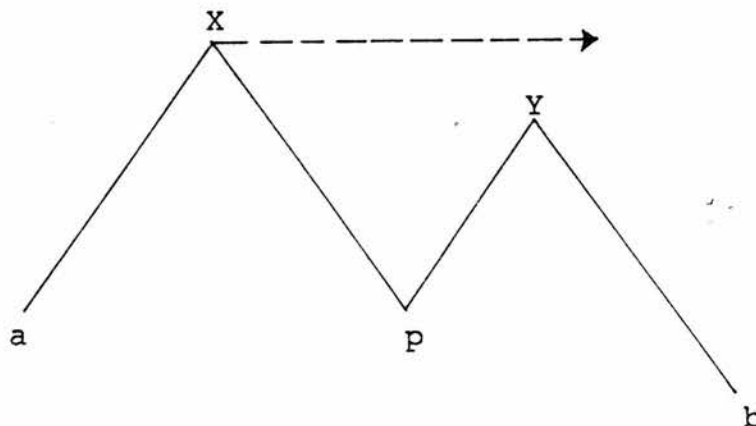
	a	X	p	Y	b
16LK and again after <u>three days</u>					
the <u>youngest</u>	190	240	190	210	160
33WM <u>I</u> will send them as brides					
for your three <u>sons</u>	120	180	120	150	90
22BS he <u>asked</u> the <u>woman</u>	110	140	110	130	100
22BS <u>tomorrow</u> <u>send</u> me your					
<u>eldest</u> son	120-180-130	160	120	140	100

In other words the fundamental frequency of X was greater than Y, but the amount of fall from both was equal. The following are examples:

	a	X	p	Y	b
16LK and again after					
<u>three days</u> the					
<u>youngest</u>		190	240	190	210 160
33WM <u>I</u> will send them as					
brides for your					
three <u>sons</u>		120	180	120	150 90
22BS he <u>asked</u> the <u>woman</u>		110	140	110	130 100
22BS <u>tomorrow</u> <u>send</u> me					
your <u>eldest</u> son	120-180-130	160	120	140	100
22BS when the widow <u>asked</u> for					
what <u>stakes</u> they					
should play	110-160-120	140	95	130	90

This contour was of the following shape; cf. Fig. 21:

Fig. 21



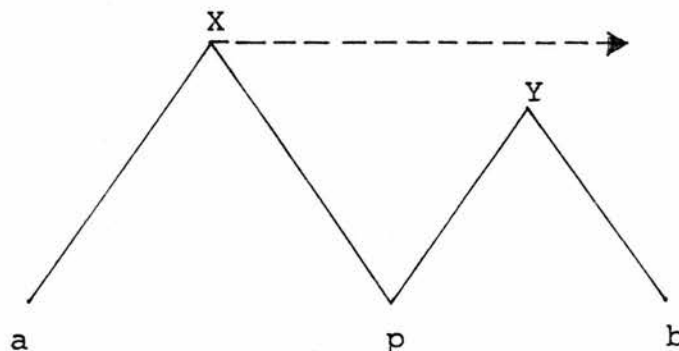
D) Reduced final fall in downward progression

These contours again involve peaks where the fundamental frequency of X is greater than the fundamental frequency of Y, but this time the amount of fall is not exactly equal in both cases, the final fall is slightly less than the initial fall, for example:

17AC when the widow <u>asked</u>	
for what <u>stakes</u> they	
should play	140 150 100 140 100
17AC the <u>next</u> day he sent	
the <u>eldest</u> son	120 150 110 135 100
16LK and the <u>old man</u>	
went <u>home</u>	160 200 160 180 150
22BS pointing <u>down</u> into	
a dark <u>valley</u>	130 140 100 130 100

The shape of the above is as follows, cf. Fig. 22:

Fig. 22



E) Reduced final fall in upward progression

These contours are similar to the ones under D) in that the final fall is less than the initial fall, but in this case the fundamental frequency of X is less than Y. There are very few examples of this type of contour, and only two are given:

17AC was a small straw hut 110-130-115 145 100 150 120

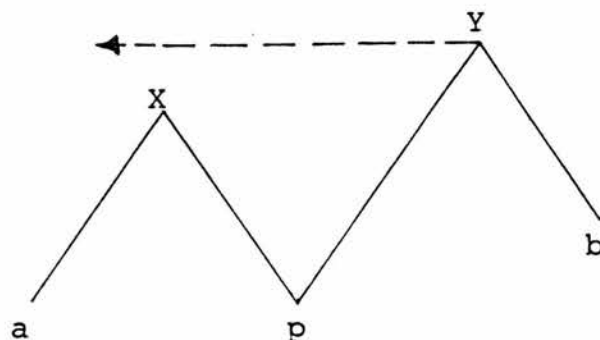
33WM without collecting

any more wood

110 140 100 155 120

This contour is of the shape shown in Fig. 23:

Fig. 23



F) Extended final fall in upward progression

This contour involves two peaks where the fundamental frequency of X is less than that of Y, and where the amount of fall from Y is greater than the amount of fall from X -- in other words the final fall is greater than the initial fall. Some examples are:

17AC that is perfect

17AC and the old man went home 105 125 95 135 95

17AC I will send them as brides

for your three sons 100 120 105 130 105

11MR and the old man went home 160 225 160 230 160

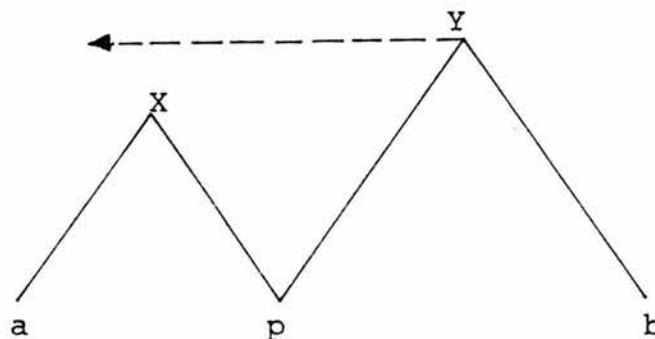
14WF you must send me your sons

as sons-in-law 200-225-180 240 170 250 160

and the shape of this contour is as follows, cf.

Fig. 24:

Fig. 24



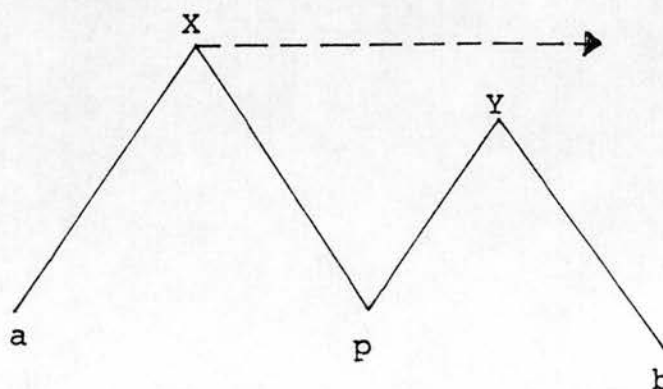
G) Extended final fall in downward progression

This contour is very similar to the one in Fig. , where the fall from Y is greater than the fall from X, but in this case the fundamental frequency of Y is less than the fundamental frequency of X. Examples of this contour are the following:

17AC	<u>tomorrow</u> send me your								
	<u>eldest</u> son	125	155	130	140	100			
33WM	<u>there</u> is my <u>house</u>	135	150	120	140	100			
33WM	three <u>sons</u> was his <u>answer</u>	160	200	150	165	95			
14WF	<u>pointing</u> down into a								
	<u>dark</u> valley	230	190	210	160				
14WF	<u>no</u> we <u>can't</u>								
	play for <u>wood</u>	190-205-180	200	170	190	160			
22BS	I will <u>send</u> them as <u>brides</u>	120	135	110	130	100			

This contour is of the following shape (cf. Fig. 25):

Fig. 25



All of the contours illustrated in figures 19-25 have been subsumed under the heading of Equal - Peaked contours because the differences between the pitch heights of the peaks, and the differences between the amounts of fall from the peaks have been very small, and on occasion, non-existent.

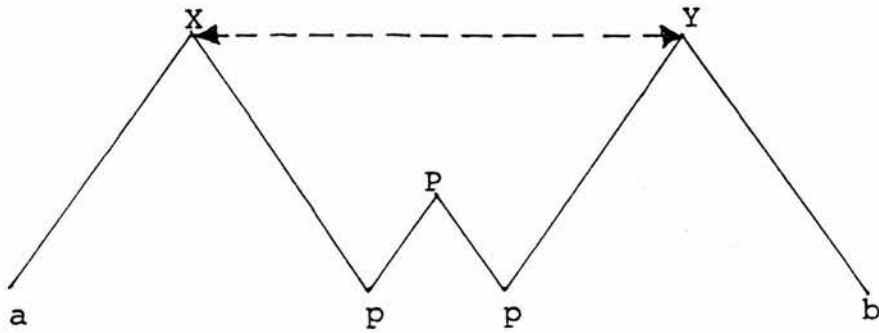
Fig. 22 is very similar to Fig. 14, but the differences in fundamental frequency between the peaks in Fig. 22 have not exceeded 10 cps, whereas in Fig. 14 a minimal difference would be 40 cps.

Similarly Fig. 24 is similar to Fig. 15, but again the maximal difference between the peaks in Fig. 24 is 10 cps. whereas in Fig. 15 the differences are much greater.

Some of the examples given in the previous pages have been contours with more than two peaks, this is because in general it was found that contours having more than two peaks could be grouped into the same sub-classifications which have been given above primarily for double-peaked contours. One additional contour must be included, however, under the heading of Equal - Peaked contours which only applies to contours which have at least 3 peaks of prominence. This contour is one where the initial peak is measured as equivalent to the final peak, but there are intervening peaks reaching only 10-30 cps above the baseline, and these intervening peaks

will be designated as pivot peaks. This contour is exemplified in Fig. 26:

Fig. 26



and examples of this contour are as follows (where the element with a broken underlining is realised as a pivot peak):

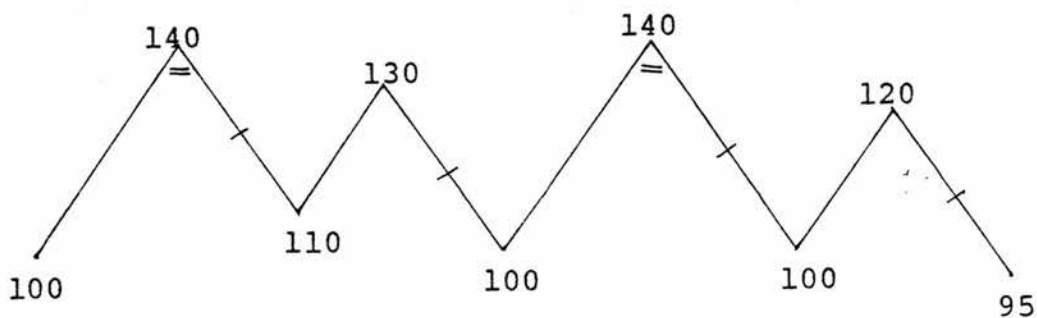
	a		X	p	P	p	Y	b
14WF <u>three sons</u> <u>that</u> is <u>perfect</u>	210	120	340	180	190	170	290	140
11MR every day the <u>father</u> went to look for <u>fuel</u>	200		225	160	180	160	200	160
16LK <u>where</u> an <u>old</u> man lived with his three <u>sons</u>	200	190	210	160	190	170	225	160

	a	X		p	P		p	Y	b		
11MR	<u>who was seated</u>										
	on a square										
	stone playing										
	<u>chess</u>	170	150	210	160	180	160	180	155	225	180

The vast majority of 3 and 4-peaked contours may be easily classified according to the sub-classification exemplified in Figs. 19-26, but because of the fact that there are more than two peaks, the relationships between 3 and 4 peaks are not so simple as the relationships between two peaks, therefore contours with more than two peaks can sometimes be classified in different ways, for example, given the following unit:

17AC	<u>who was seated</u>									
	on a square stone									
	playing chess	100	140	110	130	100	140	100	120	95

Fig. 27



In this contour $X1=Y1$, $X1p1=X2p2$, and $aX1=p2Y1$, and $Y2b=X1p1$, therefore this contour could be classified as Equal - Peaked under the headings:

- B - equal peak points
- C - equal amounts of fall
- D - reduced final fall in downward progression

The peaks are equal according to 3 different classifications, such that the contour unit would be given a complex classification instead of a simple classification because there are 4 peaks.

SUMMARY:

It will normally be quite clear which of the 3 main categories of contour the contour unit belongs to if it is one containing 2 peaks of prominence. These 3 main categories are as follows:

Contour X: exemplified in Figs. 14 and 15, is a contour which has an initial peak of prominence which is greater than any subsequent peak of prominence.

Contour Y: exemplified in Figs. 16 and 17, is a contour which has a final peak of prominence which is greater than any preceding peak of prominence.

Contour XY: exemplified in Figs. 19-26, is a contour which has two or more peaks of prominence which are equal to each other.

These three categories of contour, contour X, contour Y and contour XY, account for 226 out of the total number of 266 contour units in the six readings of the texts, the remaining 40 contours being contours with a single peak of prominence. Contours X, Y, and XY therefore, account for (approximately) 85% of the total number of contours. There are very few examples where it is difficult to assign any given unit to one of these three main categories. When these difficulties do arise the reason is normally either that the pitch movement is so small that it is difficult to know whether the peaks are equivalent or not (since the differences between the peaks are reduced anyway), or that there are 4 or 5 peaks within the contour unit, forming a complex pattern combining equal and non-equal peaks in such a way that the simple contour relationships are 'inextricably mingled'.

Contour X always occurs as the initial clause of
of a paratone.

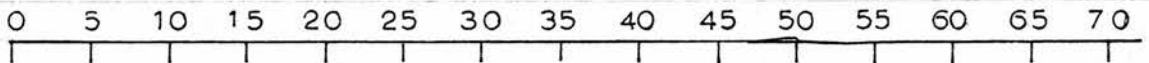
Contour Y indicates 'finality', either of turn-taking, or of a series of units.

Contour XY acts as a neutral contour unit which occurs medially within a paratone.

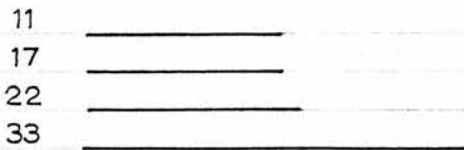
I would therefore propose that tone units (containing one stressed syllable, cf. chapter 5) are combined to form tone groups. The neutral unmarked tone group in the Edinburgh English examined for this chapter, will consist of two peaks of prominence. The relationship between the peaks of prominence within each tone group will be determined by (a) place within a paratone, (b) the structure of the surrounding discourse (producing contrastives for example, see chapter 5) and (c) the speaker's intent. The peaks of prominence will typically deviate from a baseline which consists of unstressed syllables. The tone group will typically be demarcated by boundary pauses.

Appendix 1

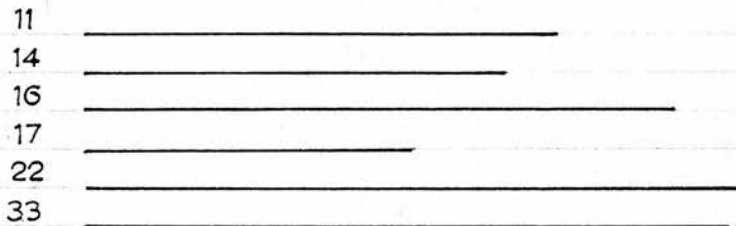
Pause duration is given in mm across the top of these graphs (1 mm = 0.02 secs). The number of each informant is given to the left of his / her pause duration. Thus informant 11 pauses for 0.25 secs after (in the midst of a range of wild mountains).



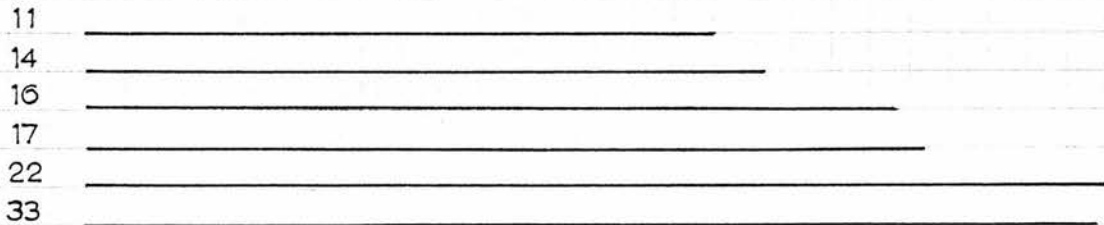
IN THE MIDST OF A RANGE OF WILD MOUNTAINS



WAS A SMALL STRAW HUT



WHERE AN OLD MAN LIVED WITH HIS THREE SONS



EVERY DAY



0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

THE FATHER WENT OUT TO LOOK FOR FUEL

11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____

ONCE HE MET IN THE WOOD AN AGED WIDOW

11 _____
 14 _____
 22 _____
 33 _____

IN WHITE CLOTHES

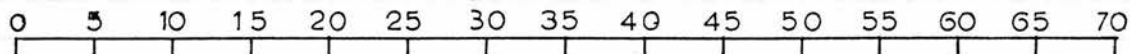
11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____

WHO WAS SEATED ON A SQUARE STONE PLAYING CHESS

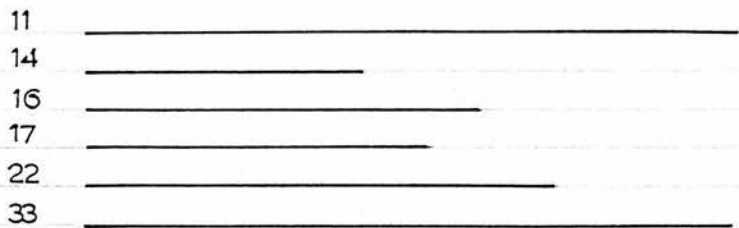
11 _____
 14 _____
 16 _____
 17 _____
 22 _____ → 77
 33 _____ → 94

SINCE THE OLD MAN WAS A KEEN PLAYER, HIMSELF

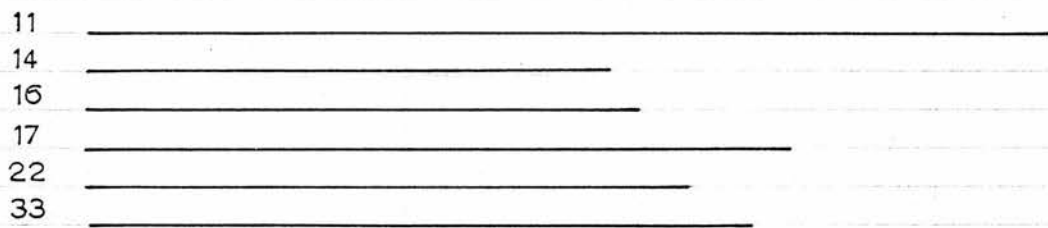
11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____



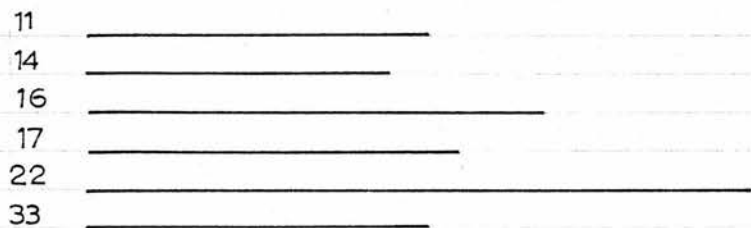
HE ASKED THE WOMAN



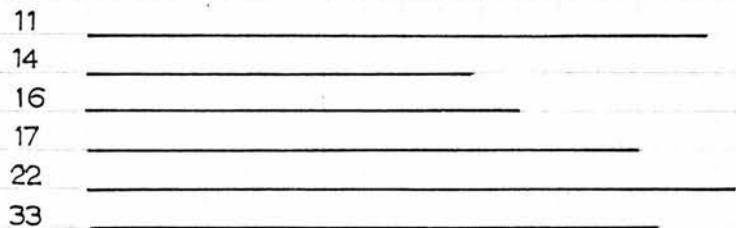
MAY I WATCH THE GAME



SHE REPLIED

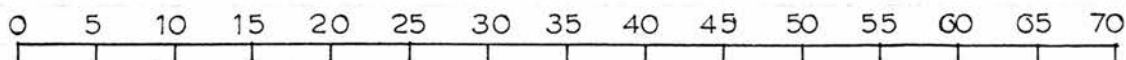


DO YOU WANT TO PLAY WITH ME

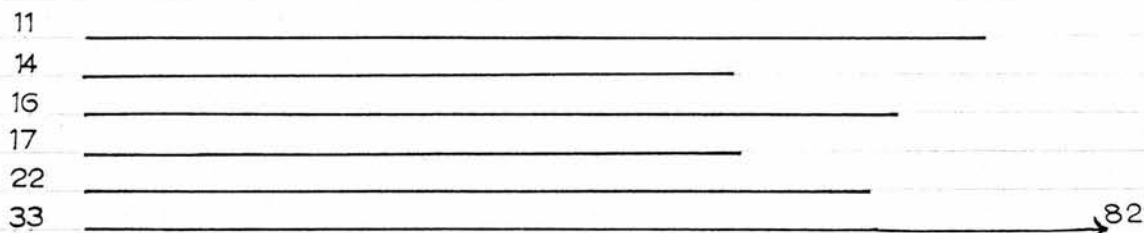


CERTAINLY

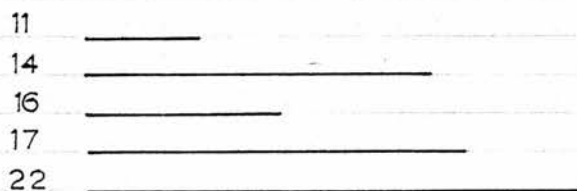




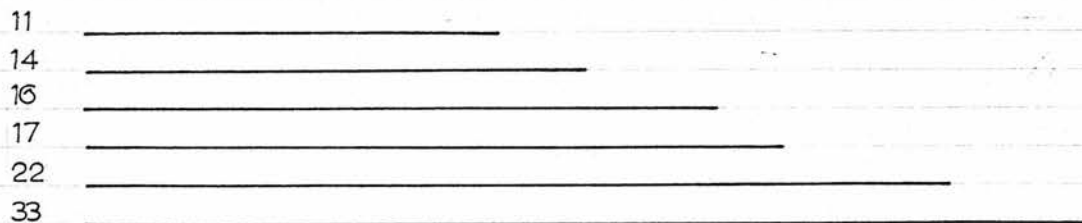
SAID THE OLD MAN



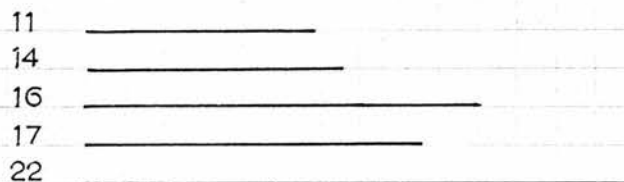
WHEN THE WIDOW ASKED FOR WHAT STAKES THEY SHOULD PLAY



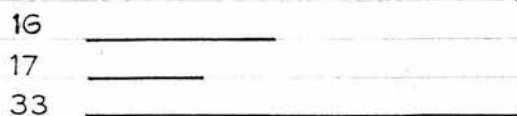
HE SUGGESTED PLAYING FOR HIS WOOD

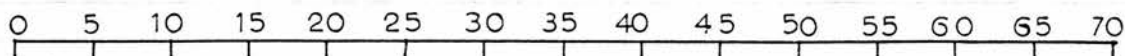


BUT THE OLD WOMAN SAID

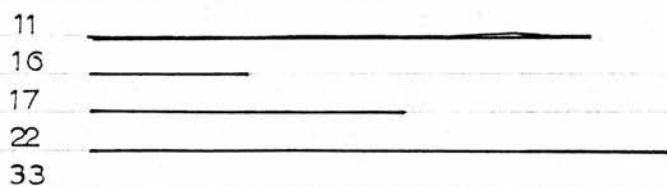


NO





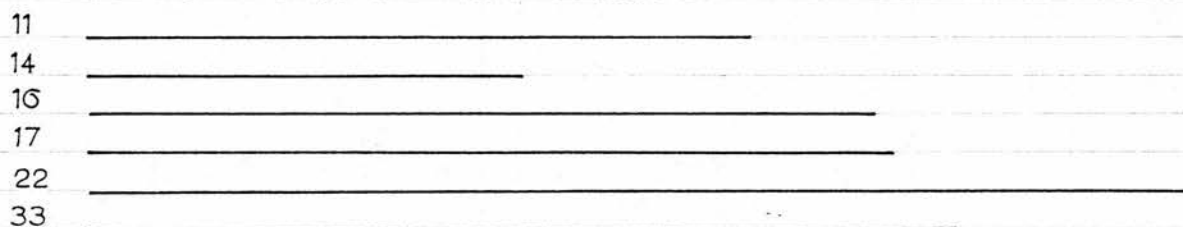
WE CAN'T PLAY FOR WOOD



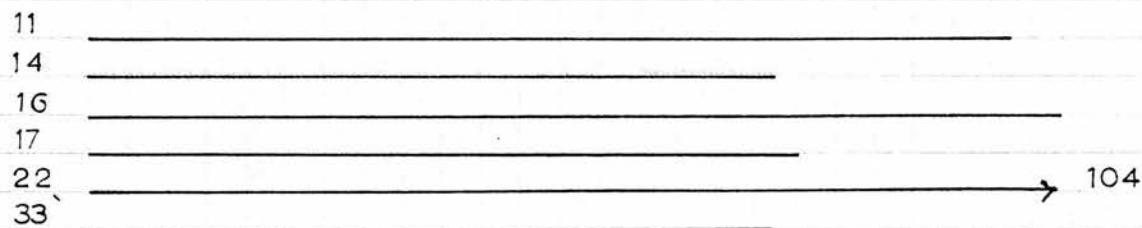
BECAUSE I



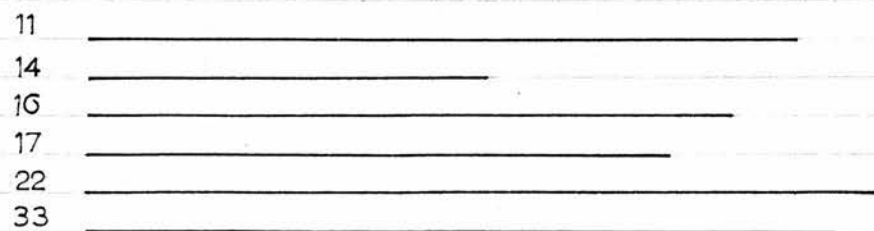
DON'T HAVE ANY WOOD

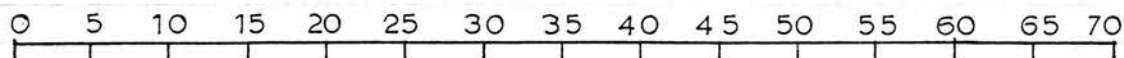


HOW MANY CHILDREN HAVE YOU THOUGH

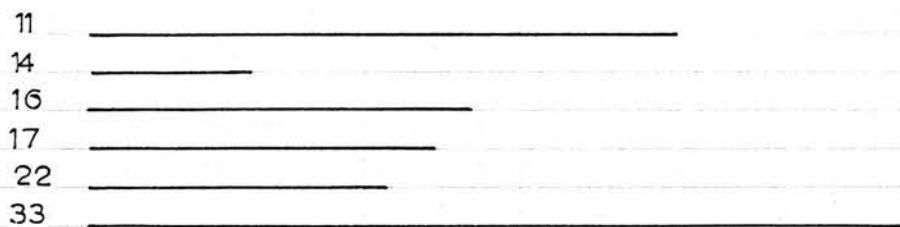


THREE SONS WAS HIS ANSWER

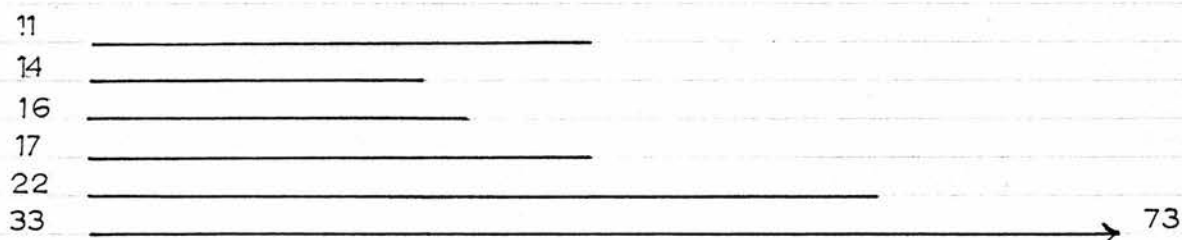




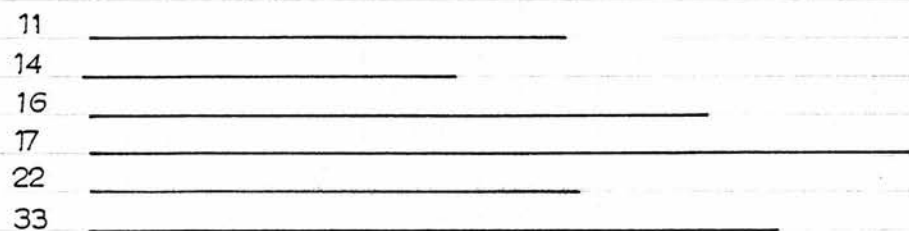
THREE SONS



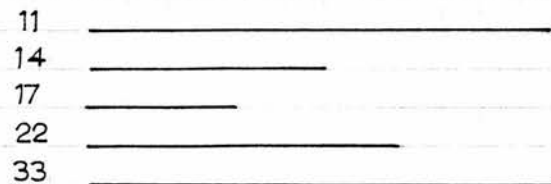
THAT IS PERFECT



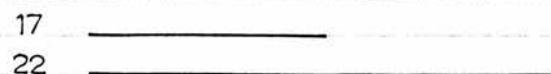
I HAVE THREE DAUGHTERS

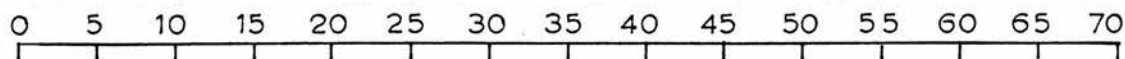


IF YOU WIN

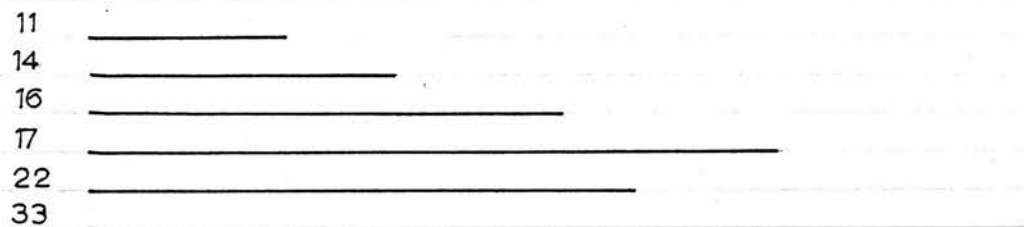


I WILL SEND THEM AS BRIDES

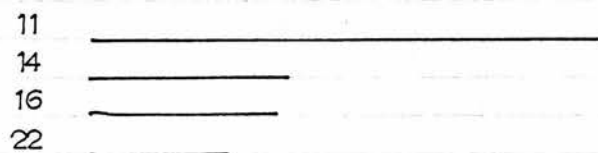




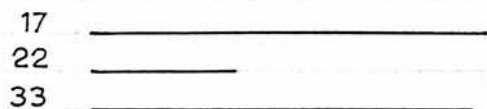
FOR YOUR THREE SONS



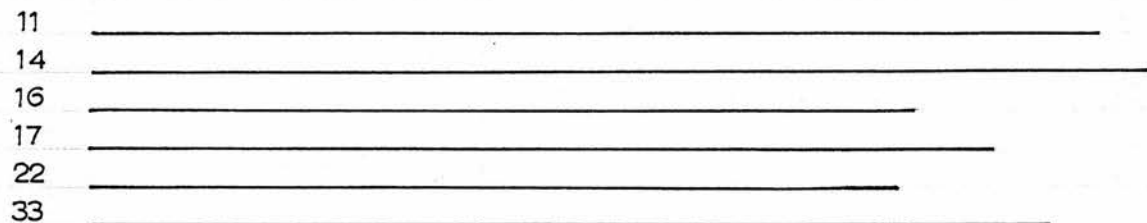
BUT IF I WIN



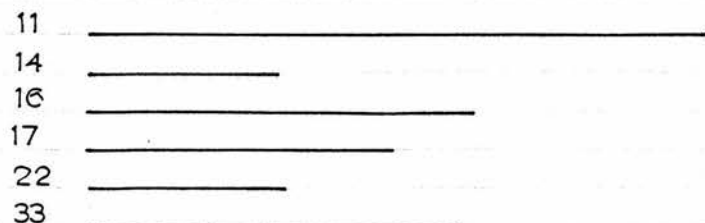
YOU MUST SEND ME YOUR SONS

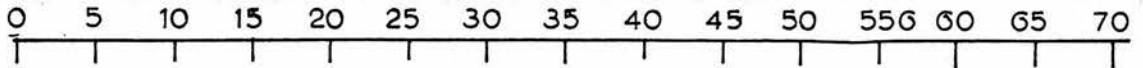


AS SONS - IN - LAW

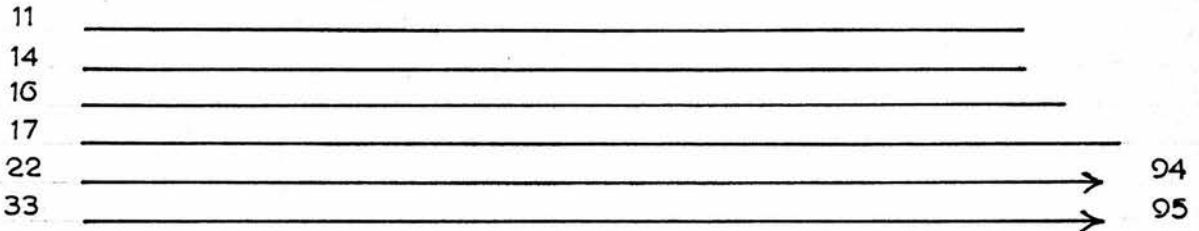


THE OLD MAN STROKED HIS BEARD FOR A WHILE

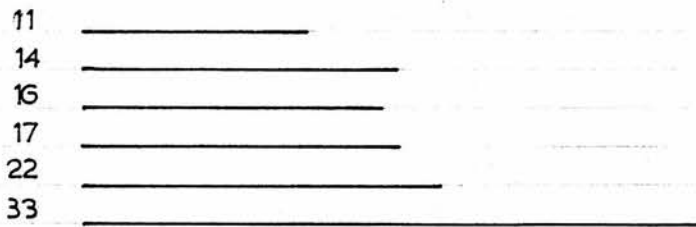




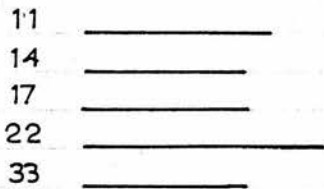
BUT FINALLY GAVE HIS ASSENT



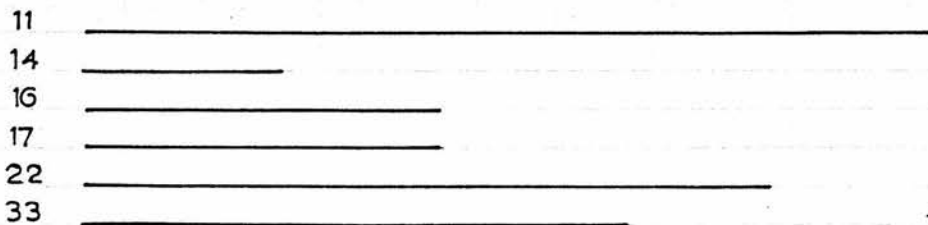
HE LOST EACH OF THE GAMES THEY PLAYED

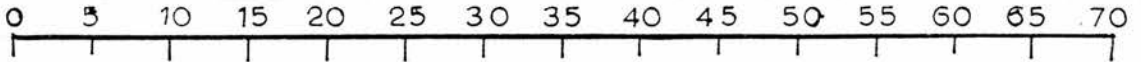


AND WHEN THE WIDOW GOT UP TO LEAVE



SHE SAID





POINTING DOWN INTO A DARK VALLEY

11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____

THERE IS MY HOUSE

11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____

TOMORROW SEND ME YOUR ELDEST SON

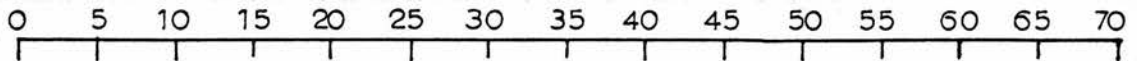
11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____ → 72

THREE DAYS LATER

22 _____

THE SECOND

11 _____
 14 _____
 16 _____
 17 _____
 22 _____
 33 _____



AND AGAIN AFTER THREE DAYS

17 _____
33 _____

THE YOUNGEST

11 _____
14 _____
16 _____
17 _____
22 _____
33 _____ → 99

SHE THEN DEPARTED

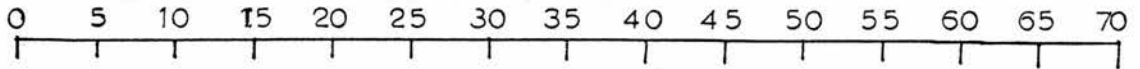
11 _____
14 _____
16 _____
17 _____
22 _____

AND THE OLD MAN WENT HOME

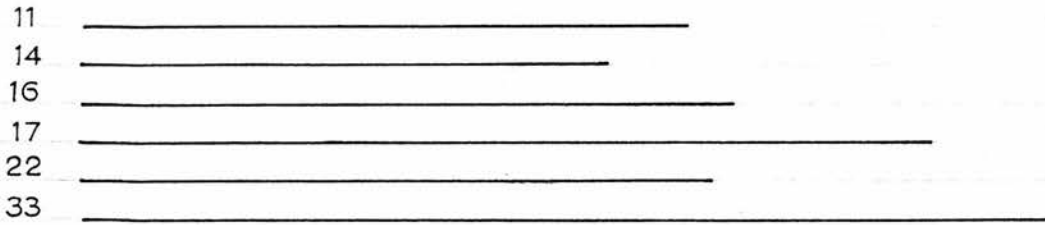
11 _____
14 _____
16 _____
17 _____
22 _____
33 _____

WITHOUT COLLECTING ANY MORE WOOD

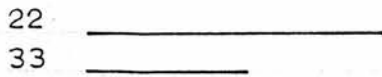
11 _____
16 _____
17 _____
22 _____
33 _____



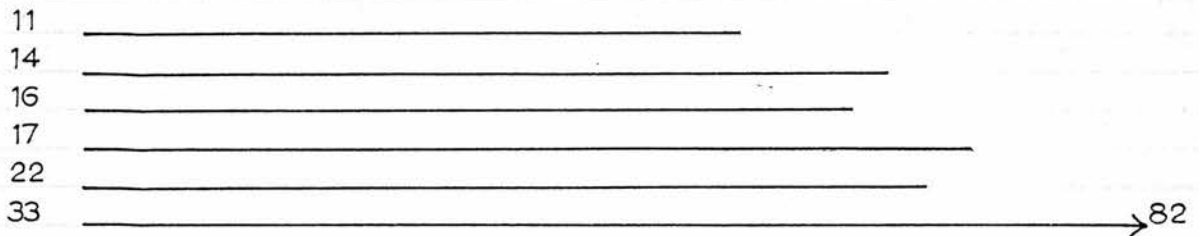
TO TELL HIS SONS WHAT HAD HAPPENED



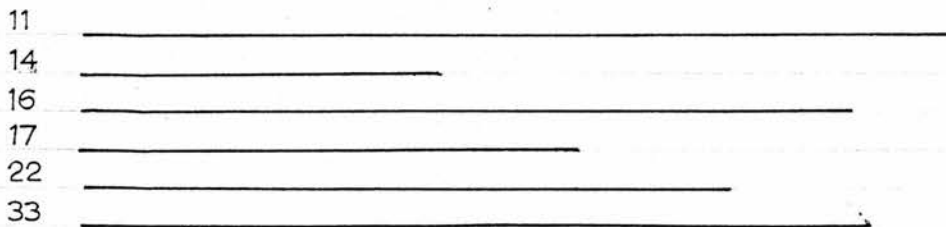
HOW PLEASED THEY WERE

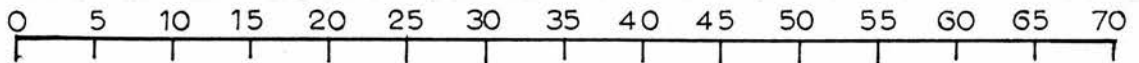


WHEN THEY HEARD IT

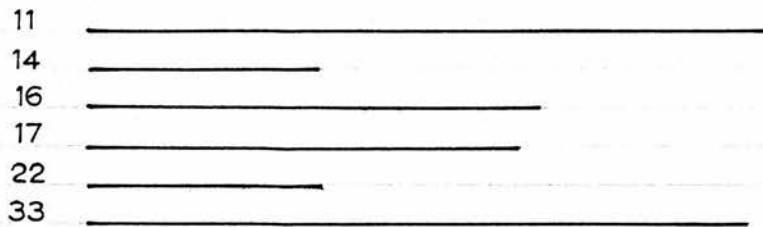


THE NEXT DAY HE SENT THE ELDEST SON

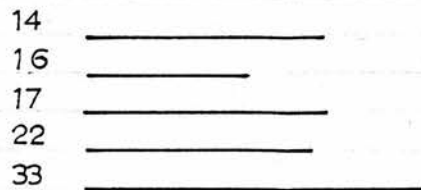




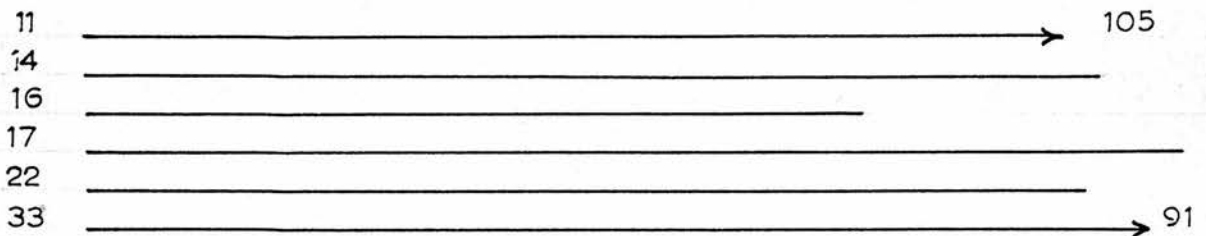
THREE DAYS LATER THE SECOND



AND ON THE SIXTH DAY



HE SENT THE YOUNGEST



(PRECEDING INTERRUPTION BY INTERVIEWER)

Appendix 2

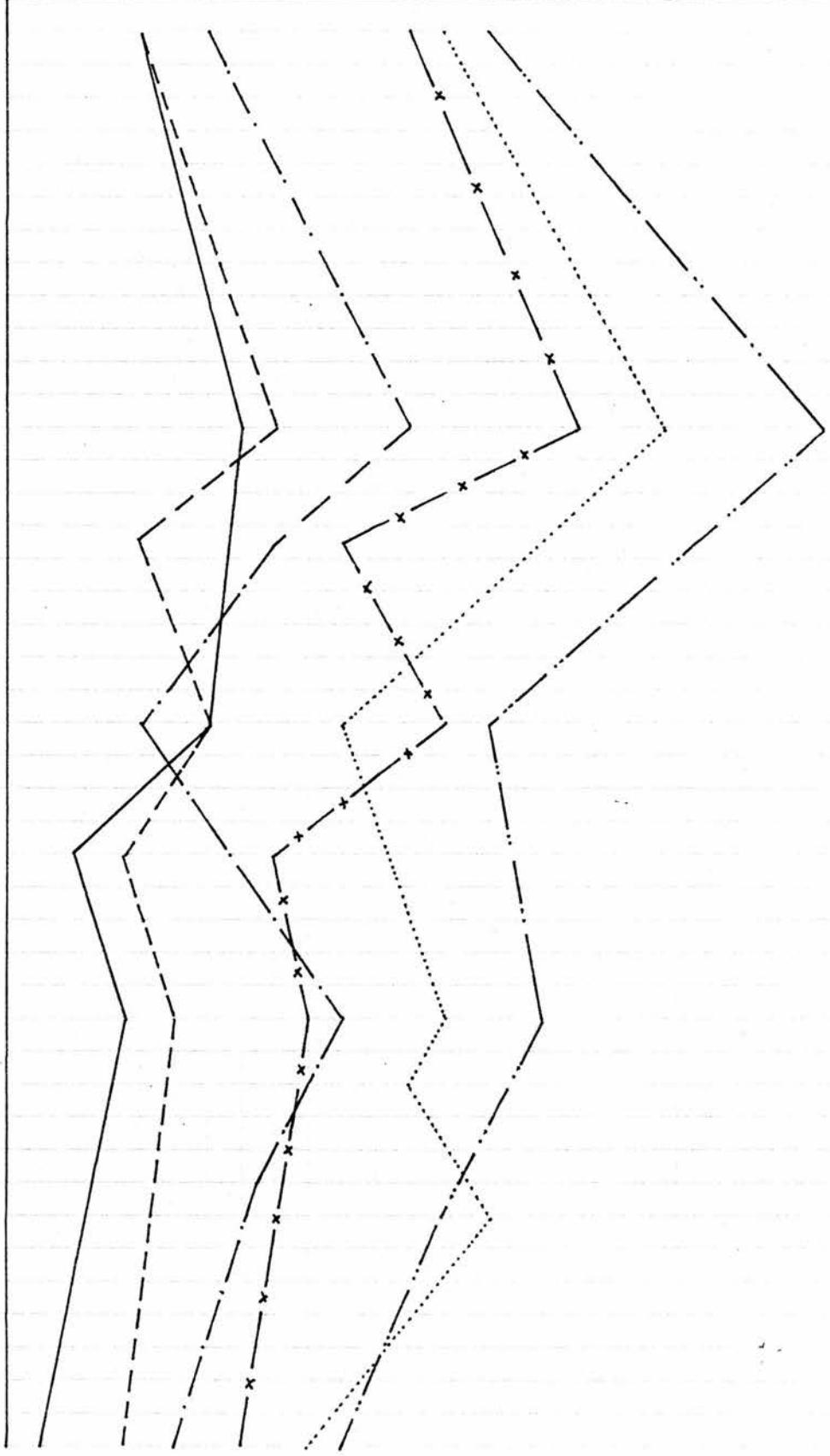
In the following graphs, fundamental frequency is given along the top axis in cycles per second. The text is given above the appropriate section of the contour. Words which are written in capitals are realised as peaks of prominence. Words which are underlined are realised as peaks of prominence by at least 5 out of the 6 speakers, words which are surrounded by brackets are realised as peaks of prominence by 4 or less speakers. The graphs of each speaker are represented as follows:

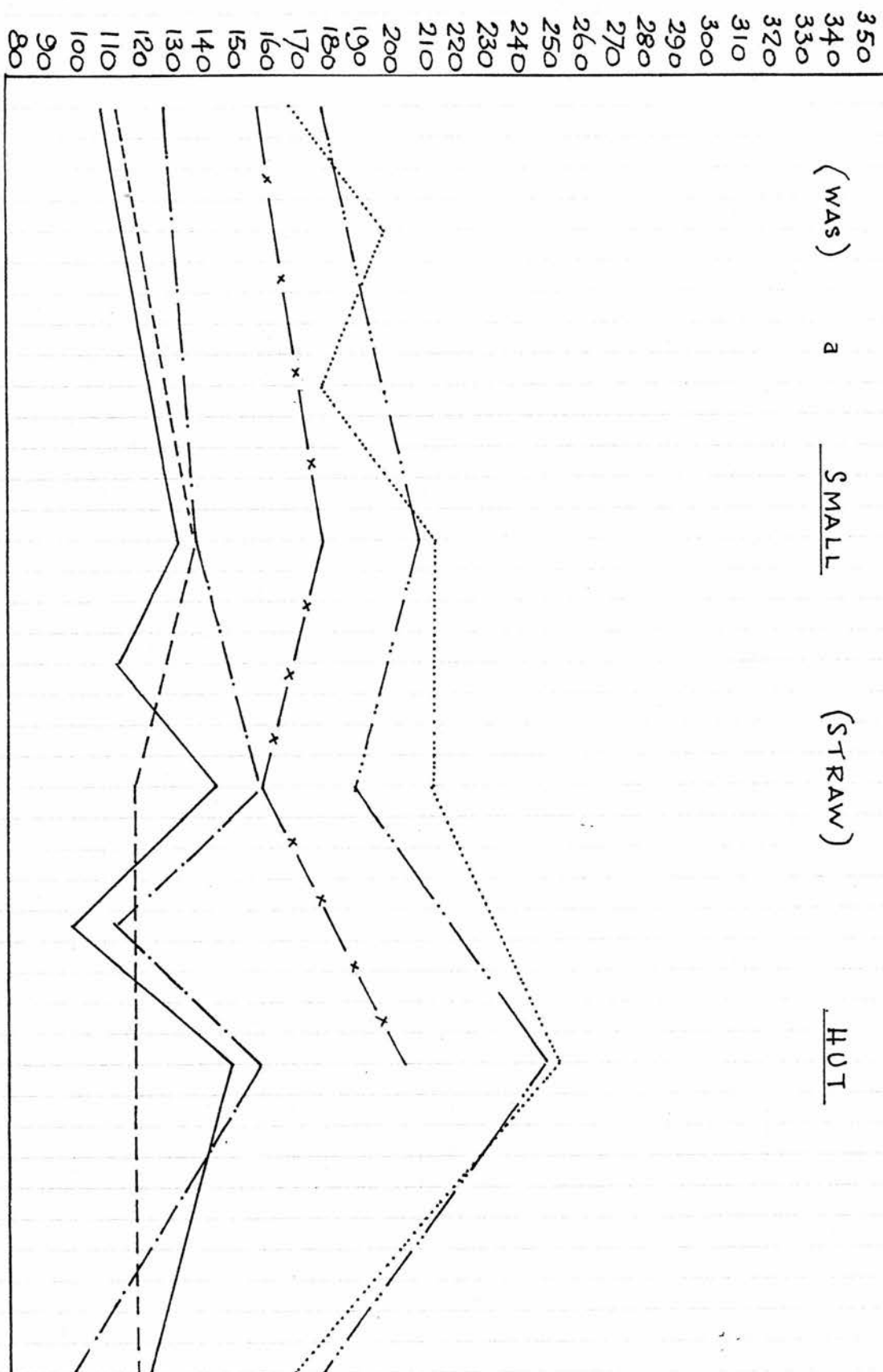
11MR	——— X ——— X ———
14WF	——— .. ——— .. ———
16LK
17AC	—————
22BS	-----
33WM	—— . —— . —— . ——

These graphs are a diagrammatic representation of the measurement of the 'extreme' points of pitch movement (i.e. the highest or lowest point preceding a change of pitch direction was measured). These graphs make no claims about the realisation of pitch movement between these extremes.

in the MIDST of a (RANGE) of WILD (MOUNTAINS)

350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80



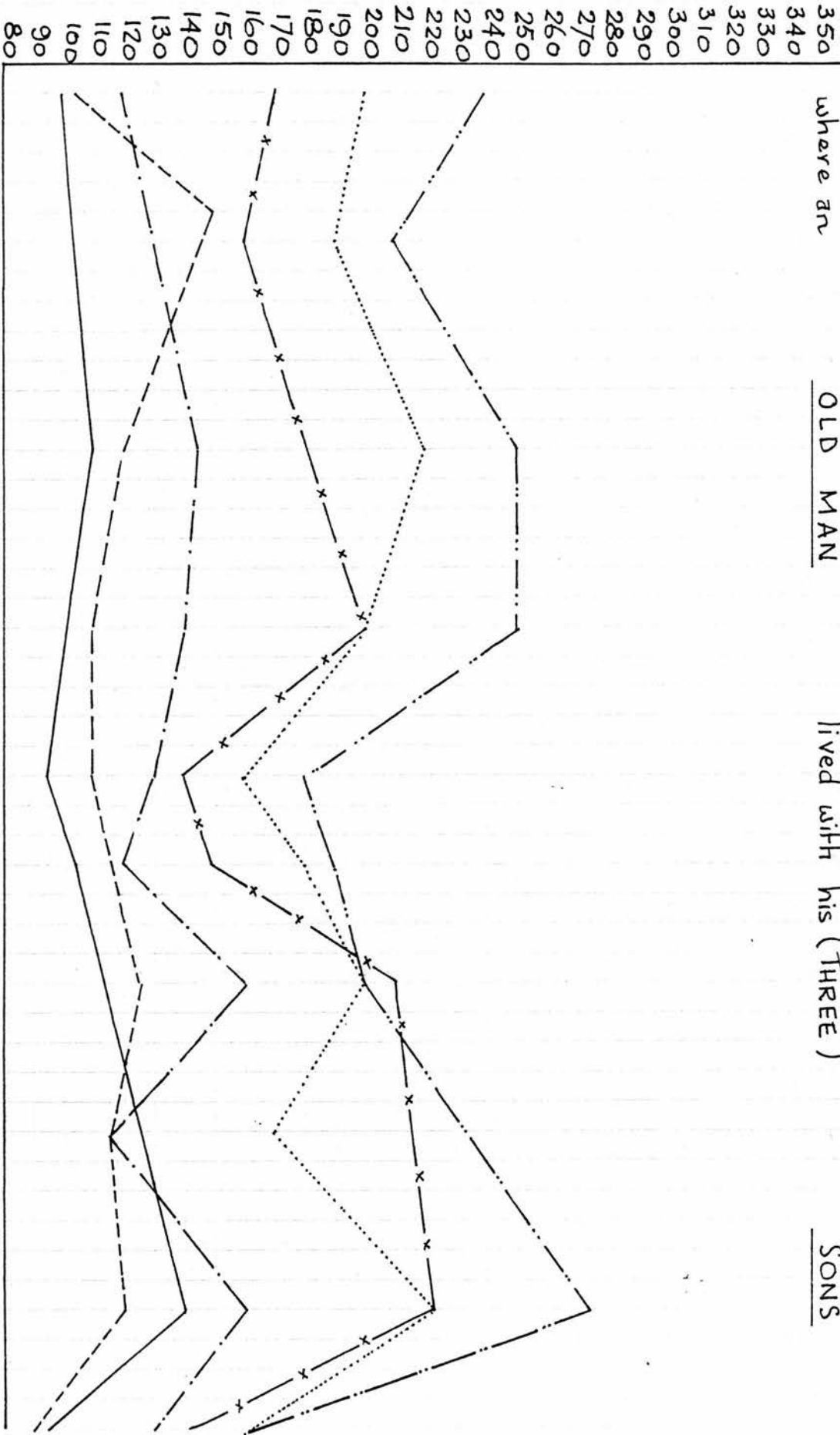


where an

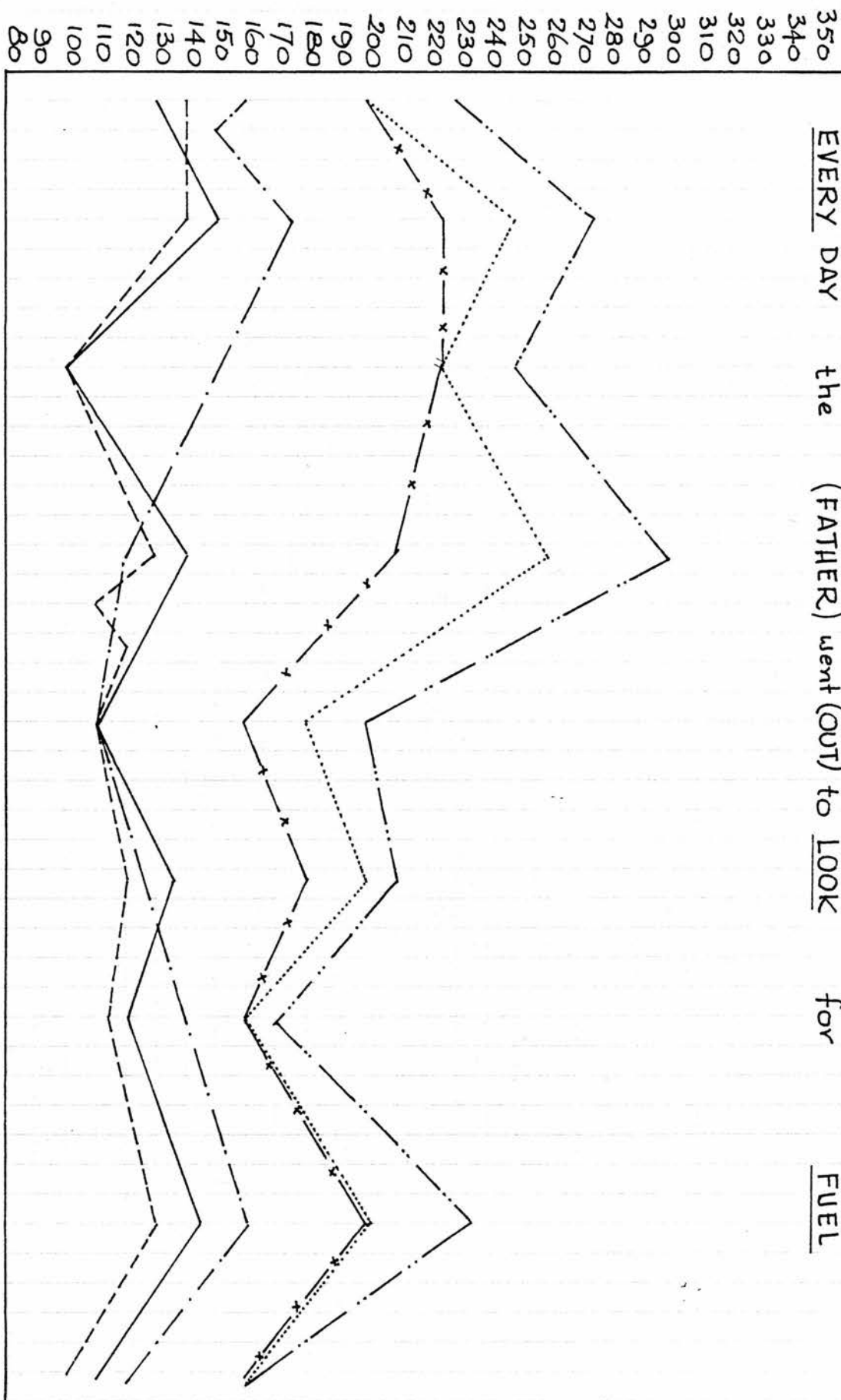
OLD MAN

lived with his (THREE)

SONS



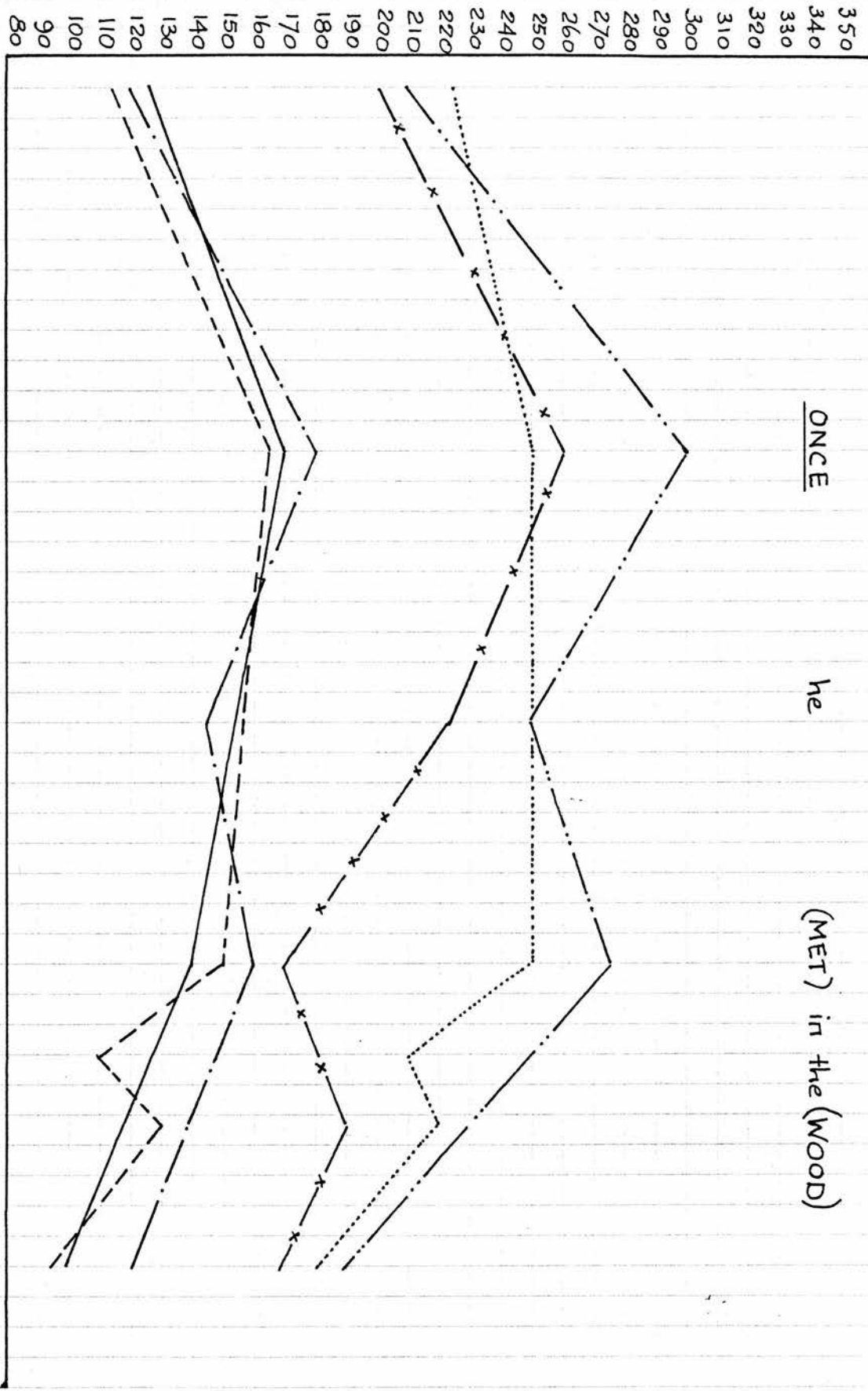
EVERY DAY the (FATHER) went (OUT) to LOOK for FUEL

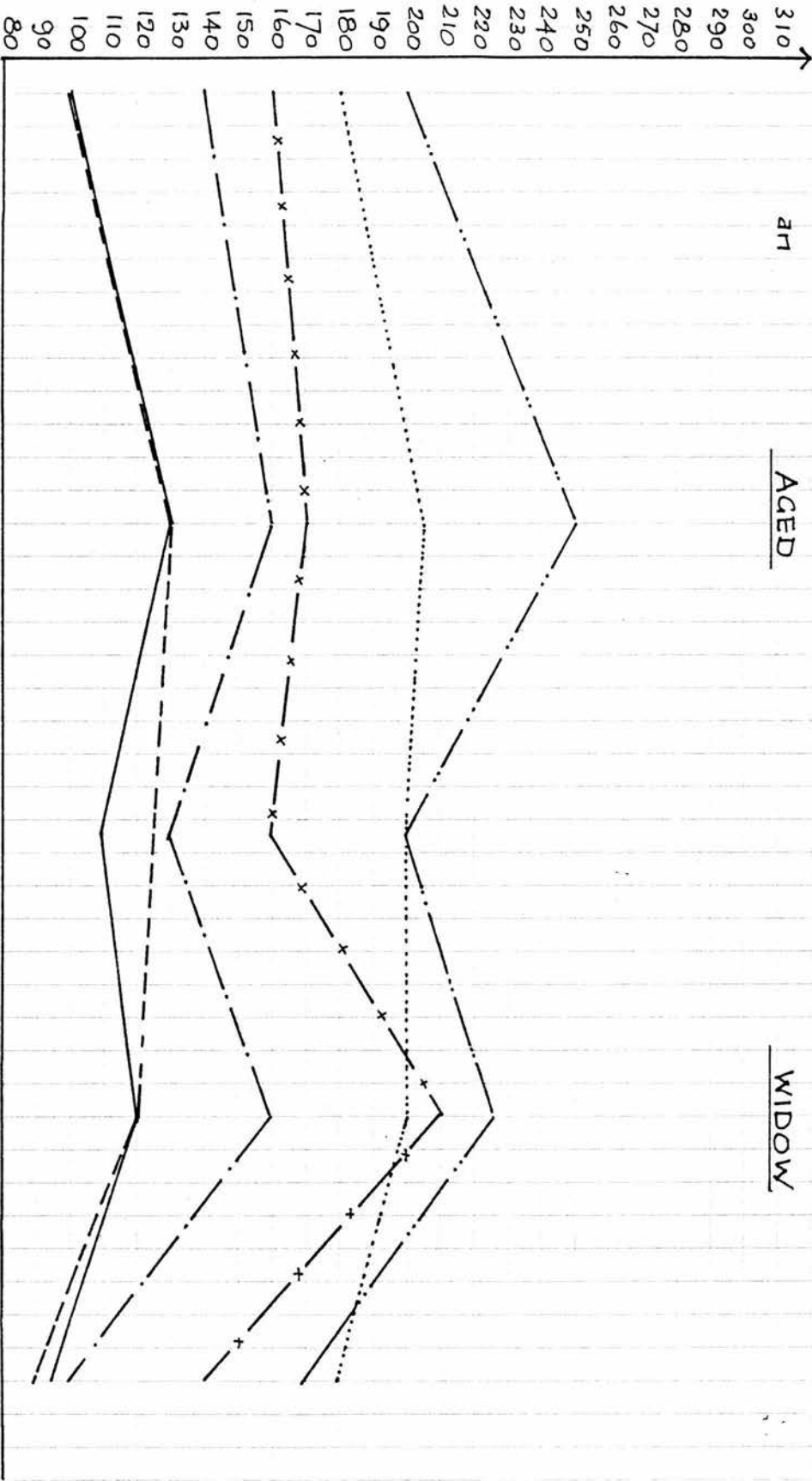


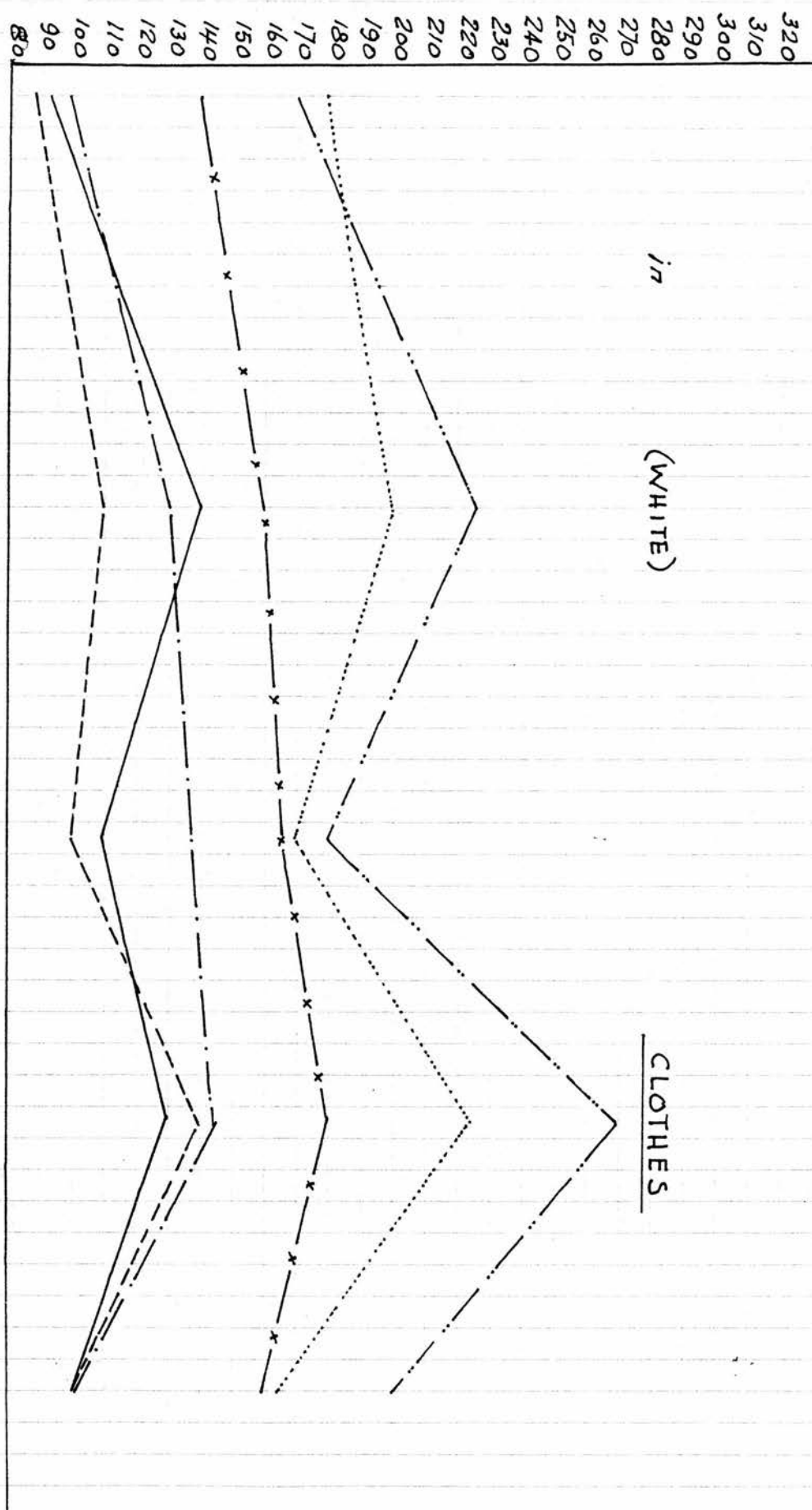
ONCE

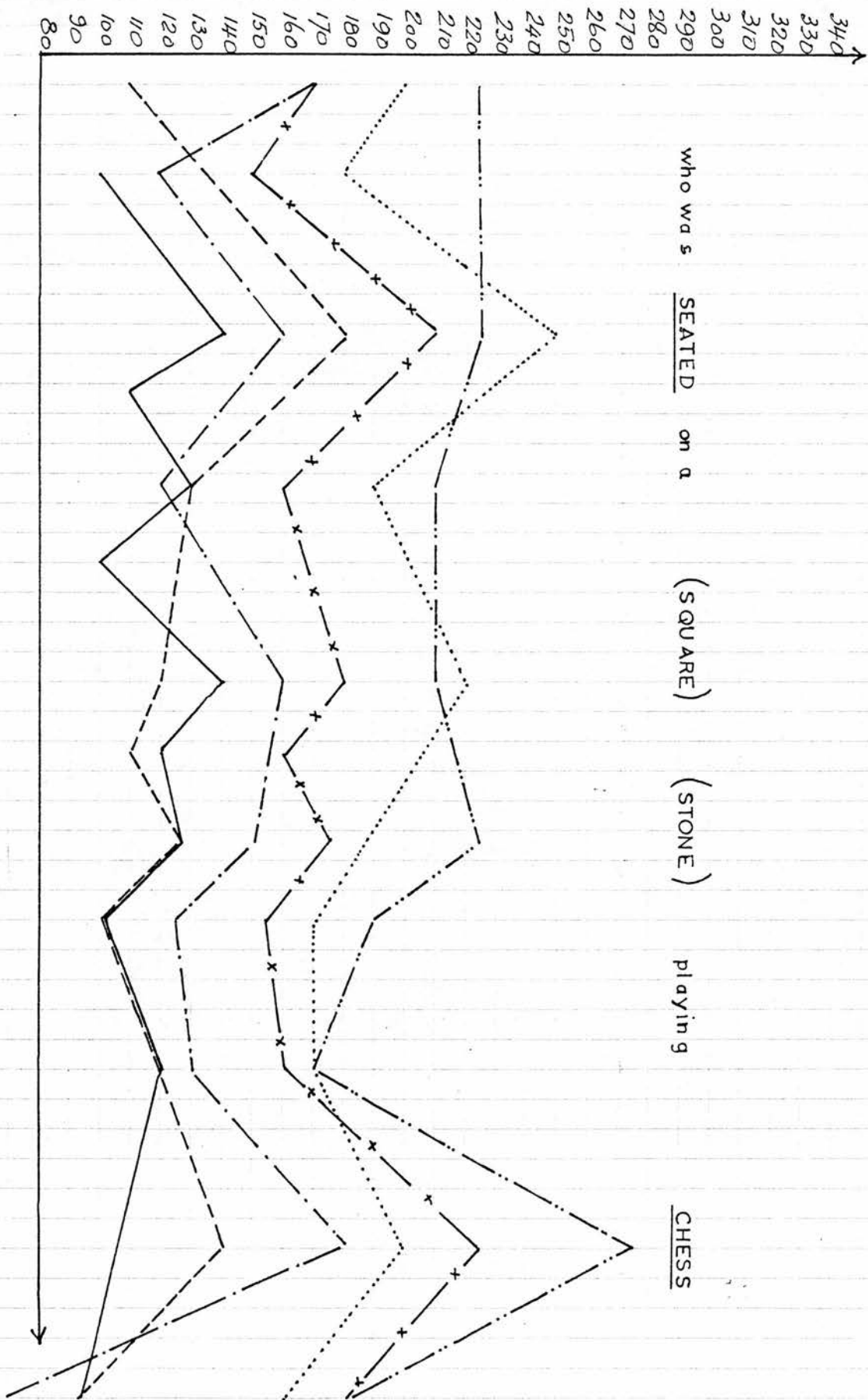
he

(MET) in the (WOOD)



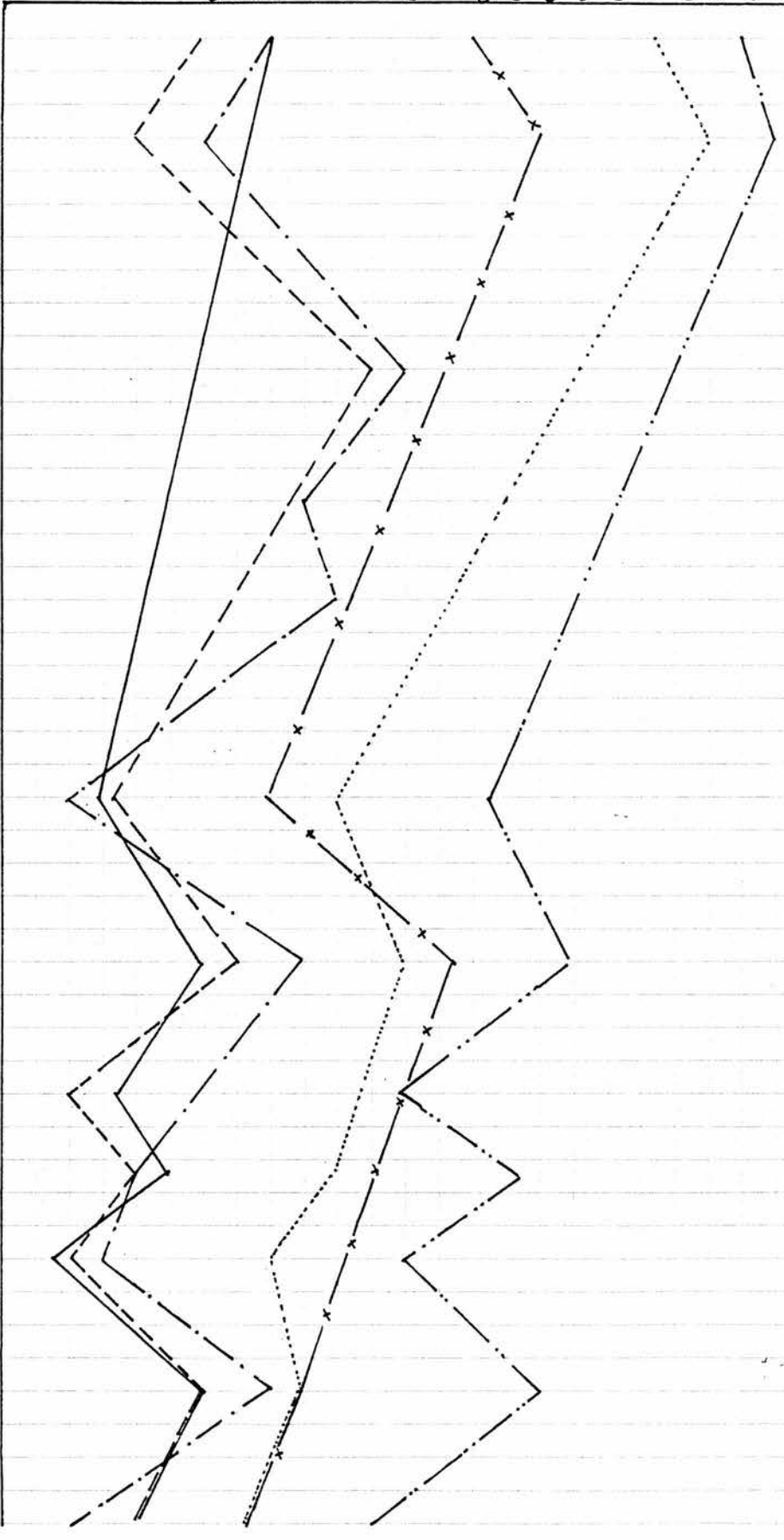






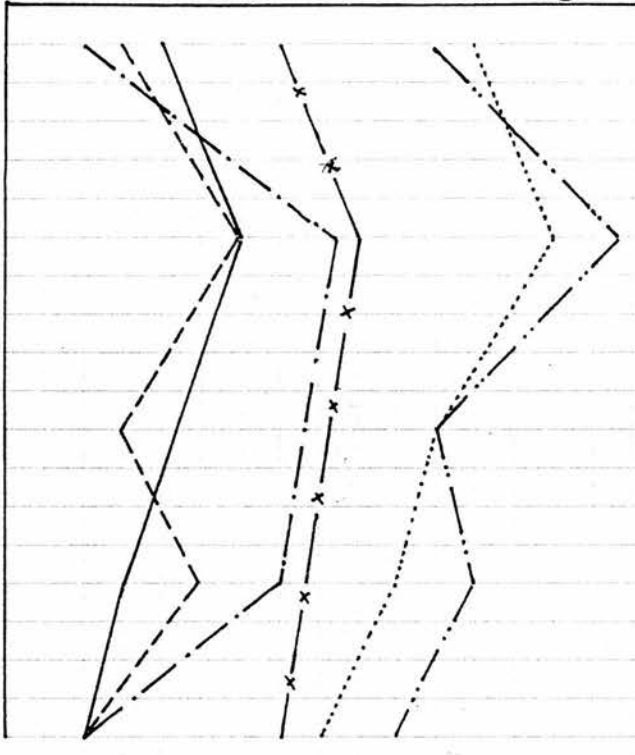
SINCE he (OLD) (MAN) was a KEEN (PLAYER) HIMSELF

350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80

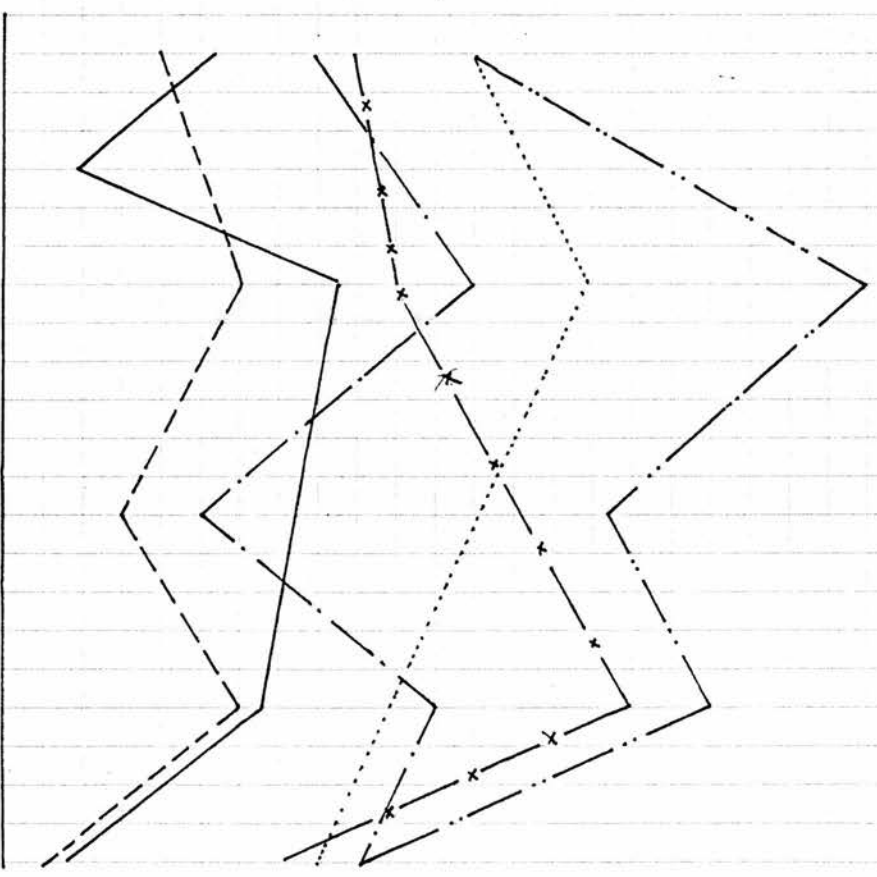


350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80

he ASKED the (WOMAN)



may I WATCH the (GAME)

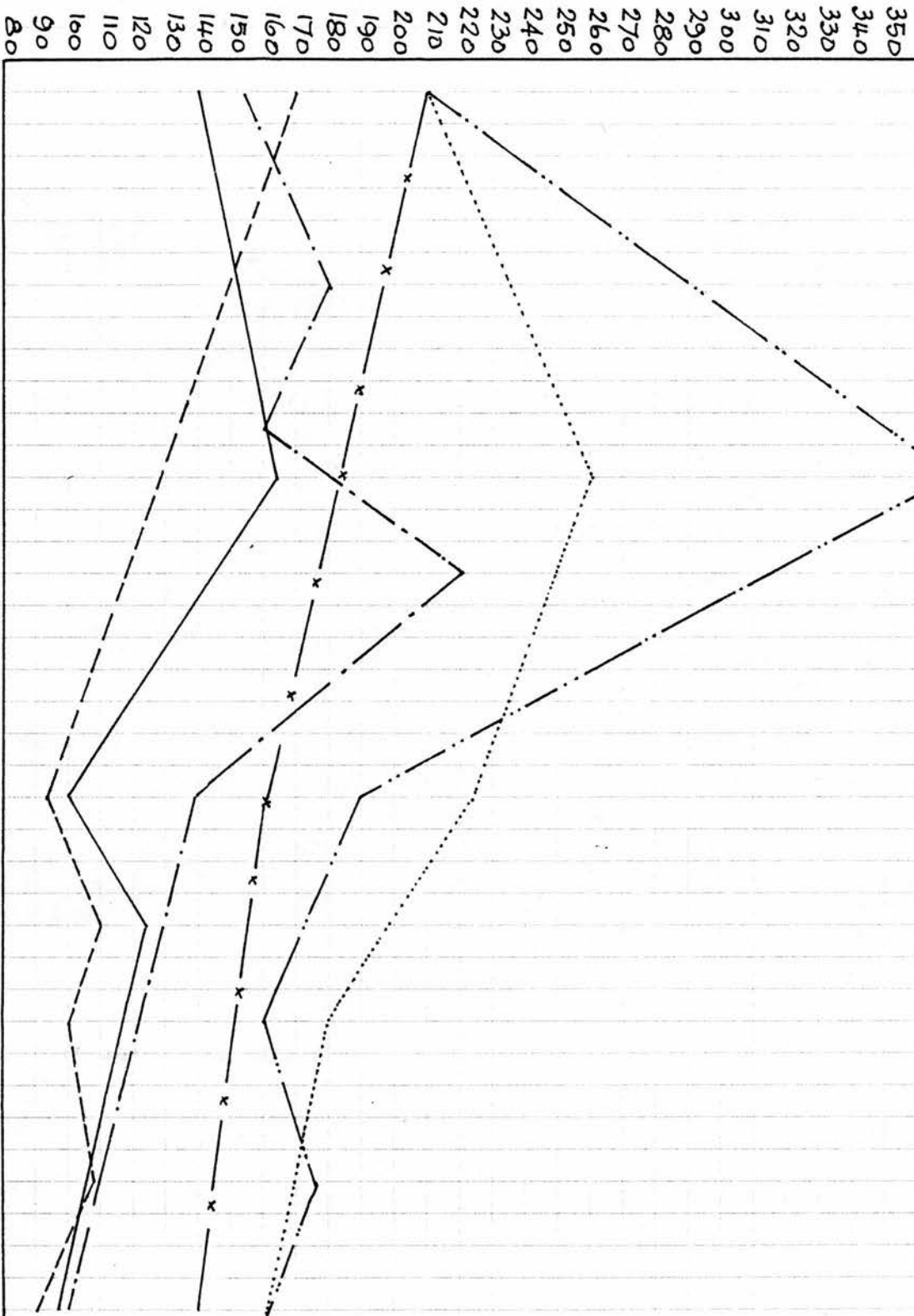


CERTAINLY

said the

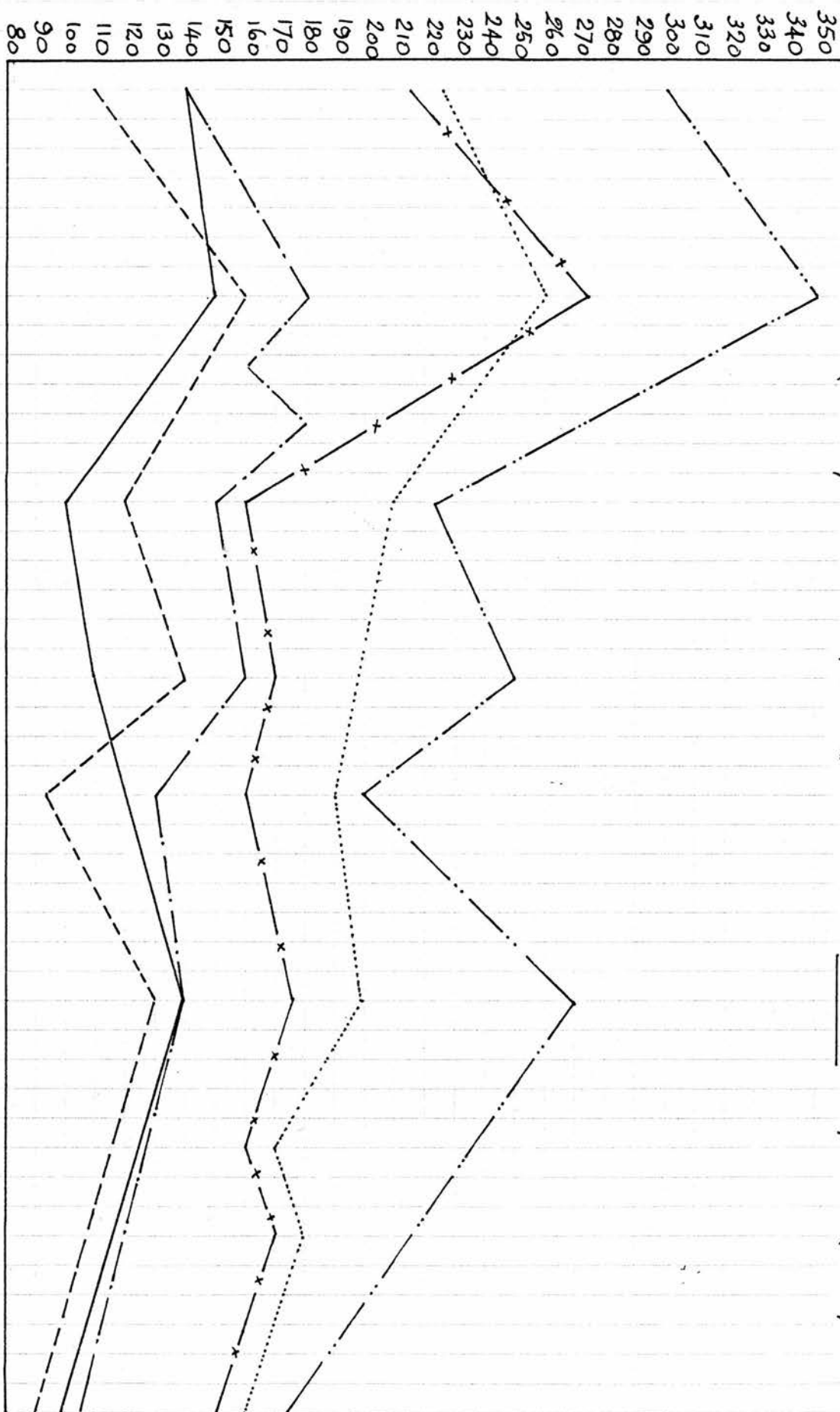
(OLD)

(MAN)



when the WIDOW (ASKED) for (WHAT)

STAKES they should (PLAY)

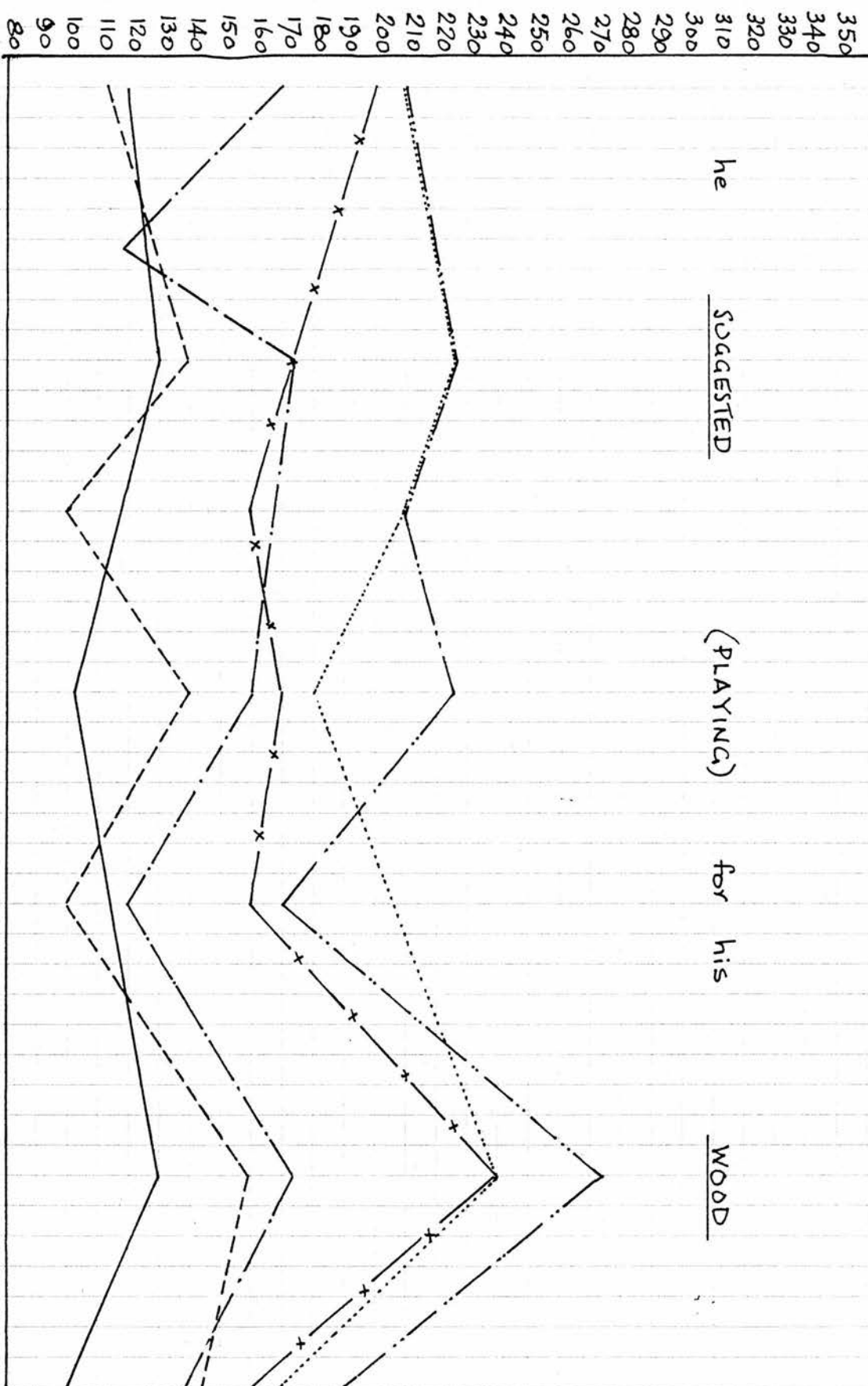


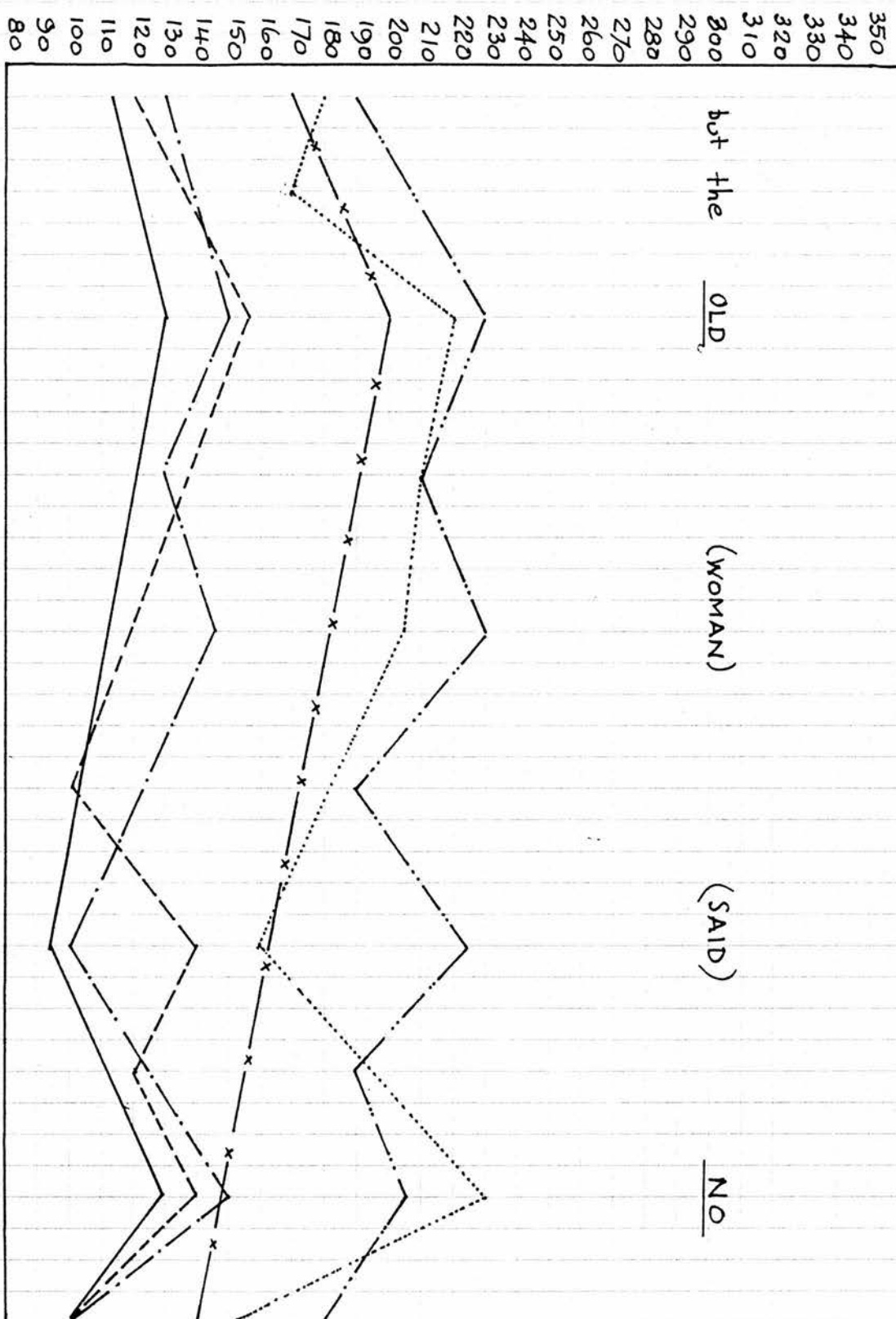
he

SUGGESTED

(PLAYING)

for his

WOOD



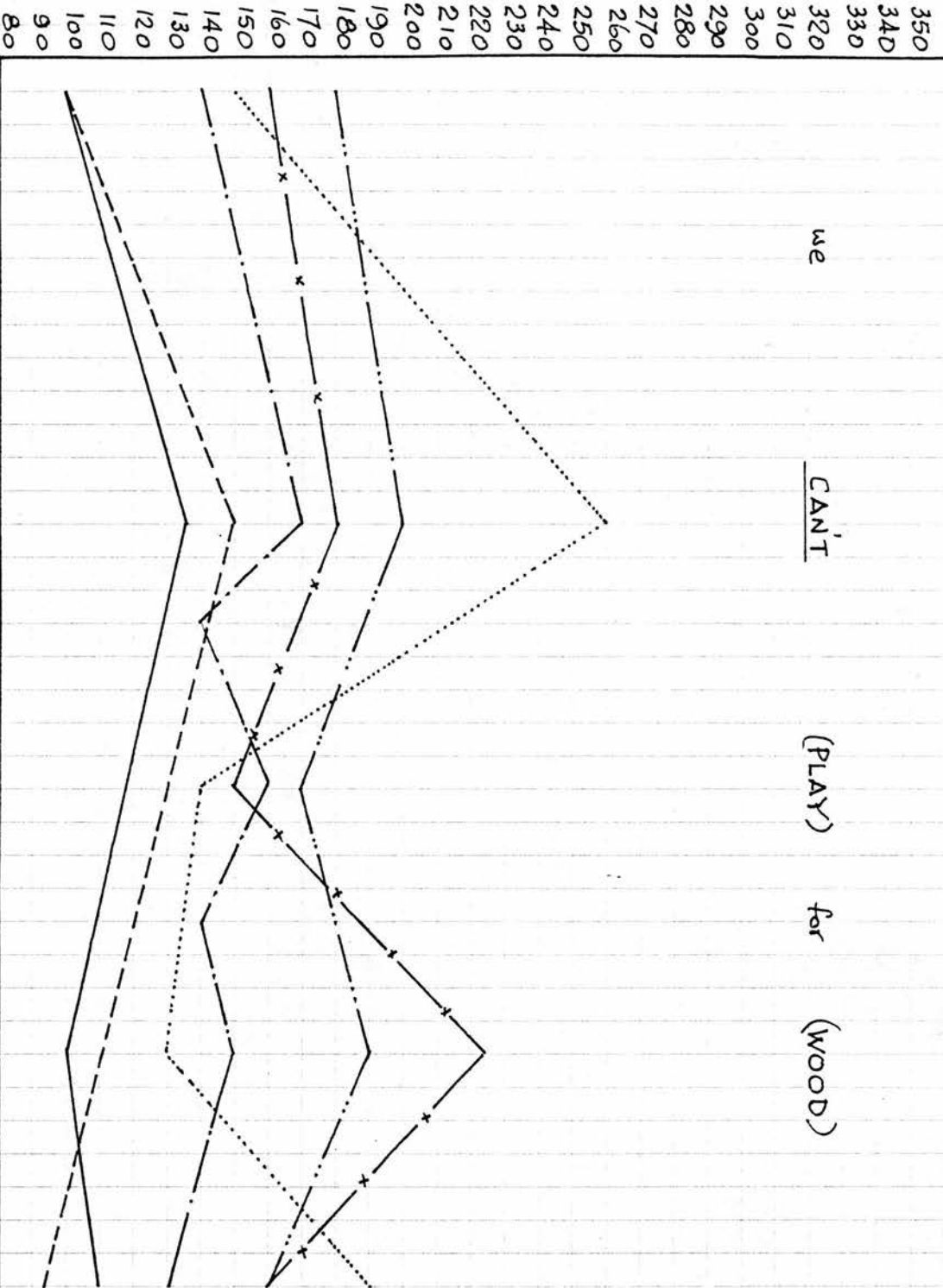
we

CAN'T

(PLAY)

for

(WOOD)



350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80

because

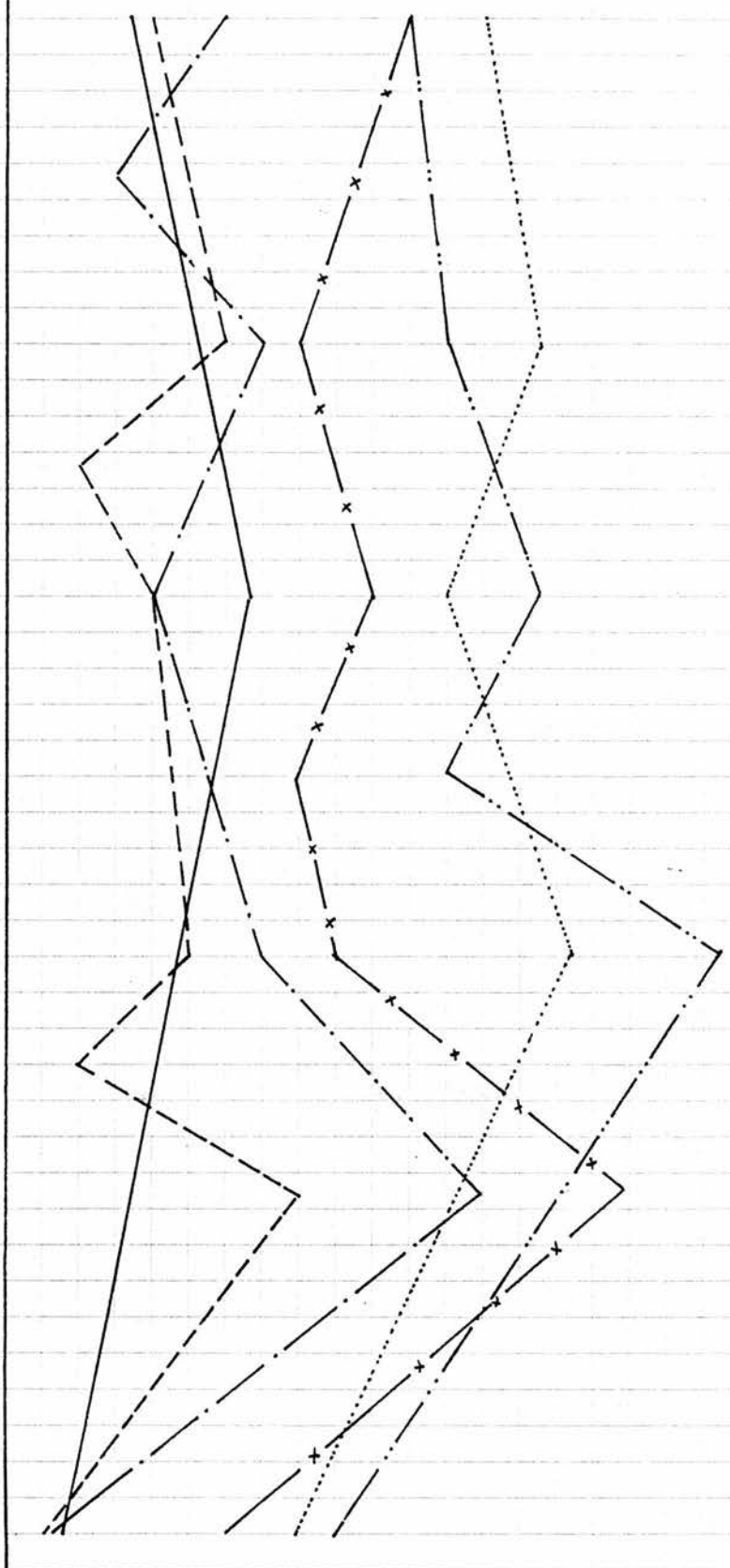
(I)

(DON'T)

(HAVE)

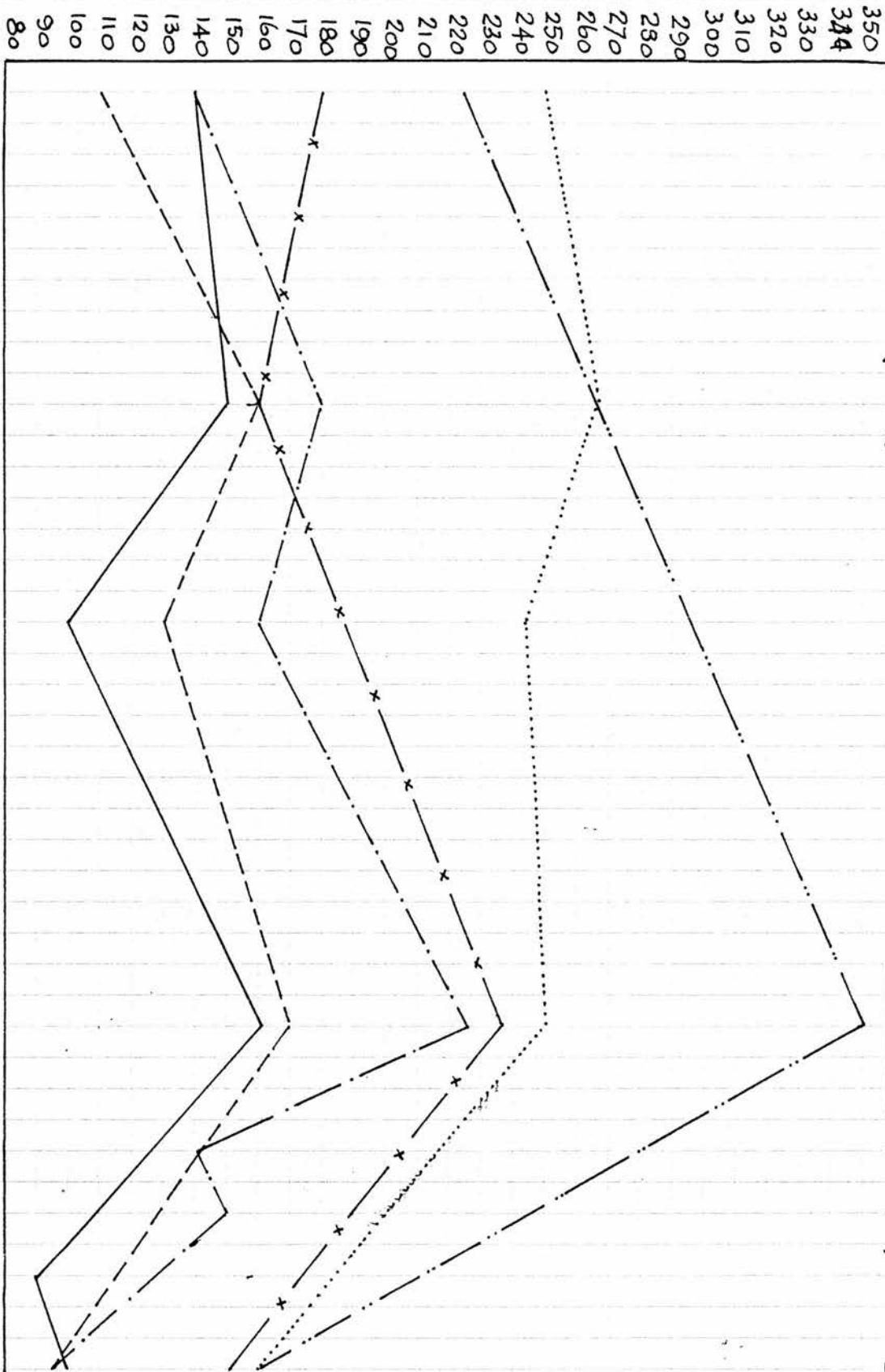
(ANY)

wood



how (MANY)

CHILDREN (HAVE) you though?

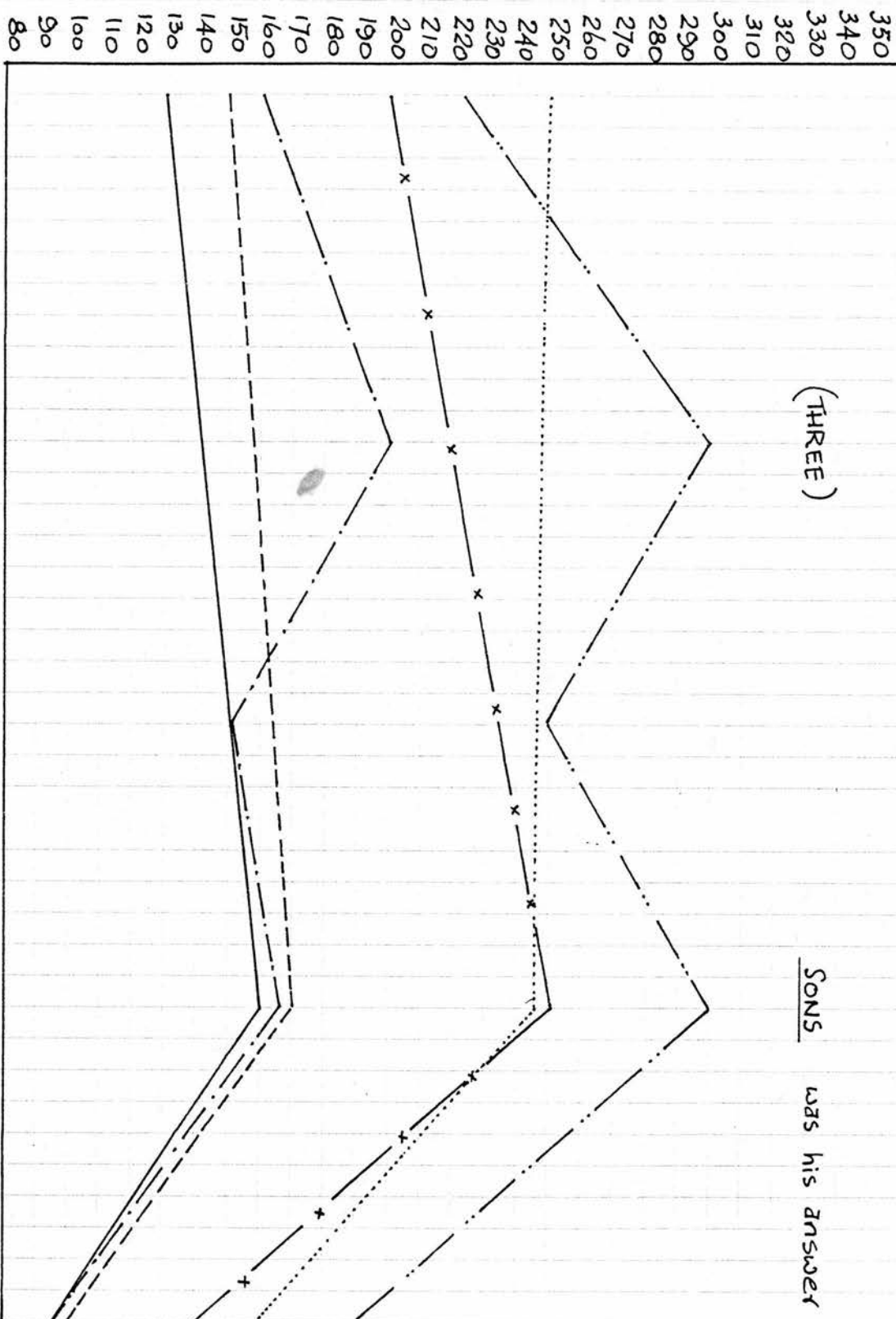


350
344
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80

how (MANY)

CHILDREN (HAVE) you though?

(THREE)

SONS was his answer

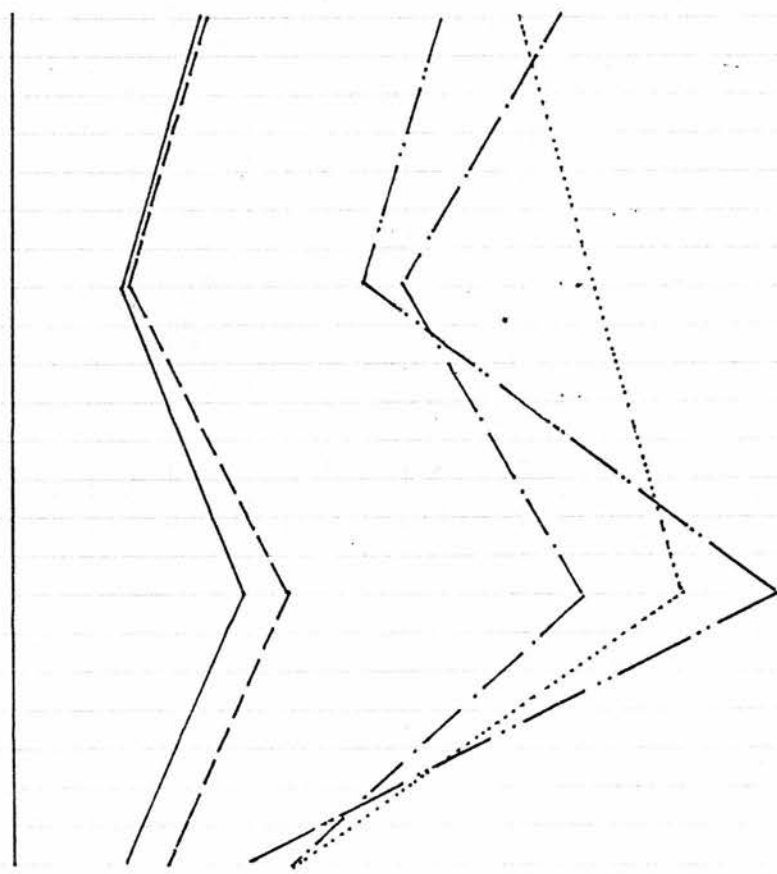
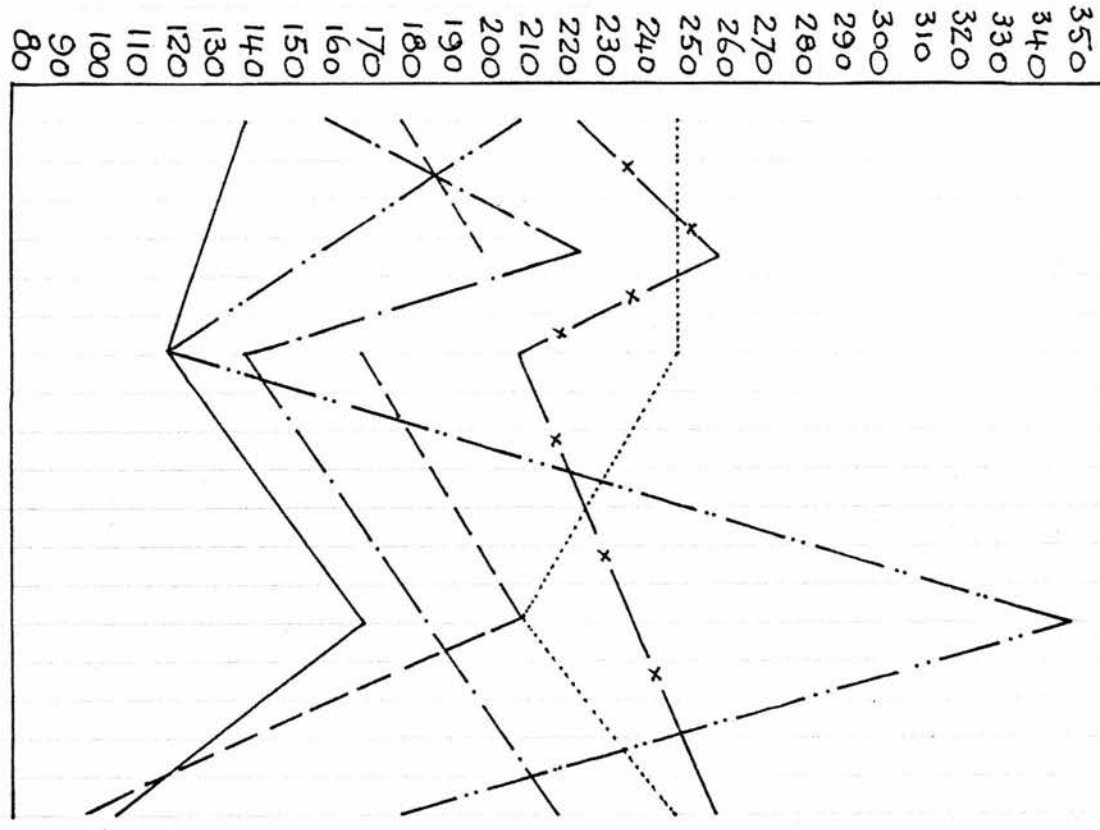
(THREE)

(SONS)

THAT

IS

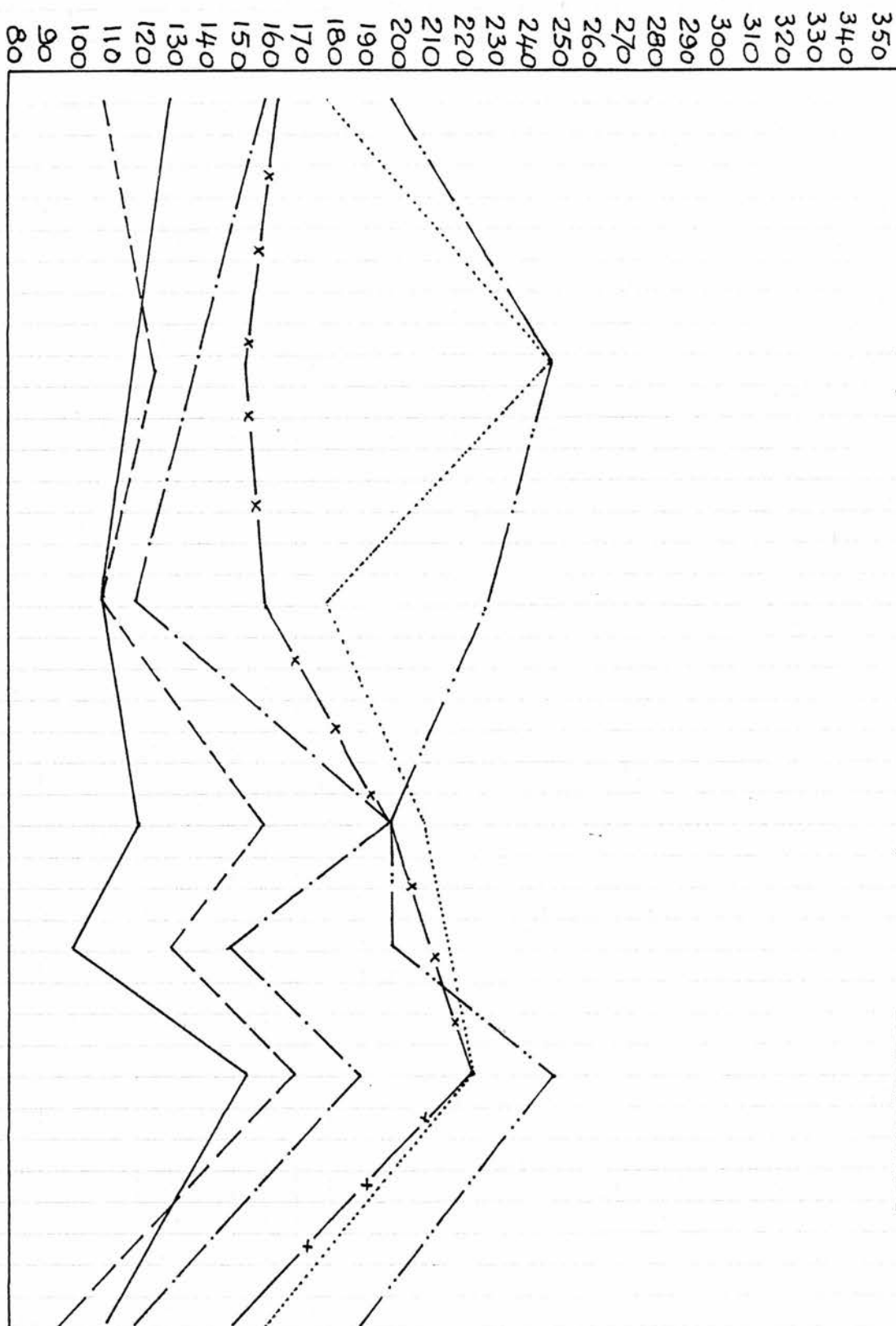
PERFECT



(I)

have

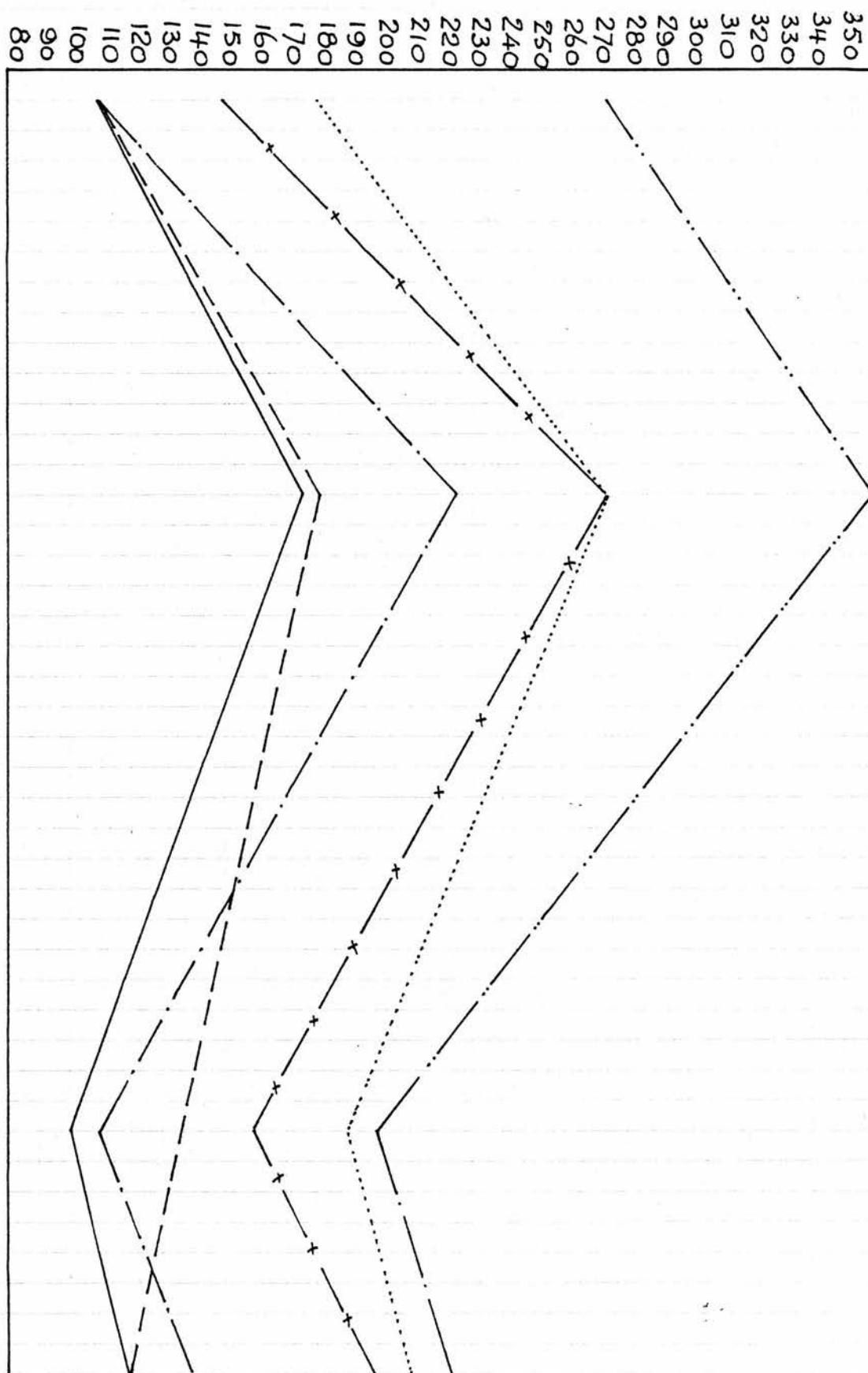
(THREE)

DAUGHTERS

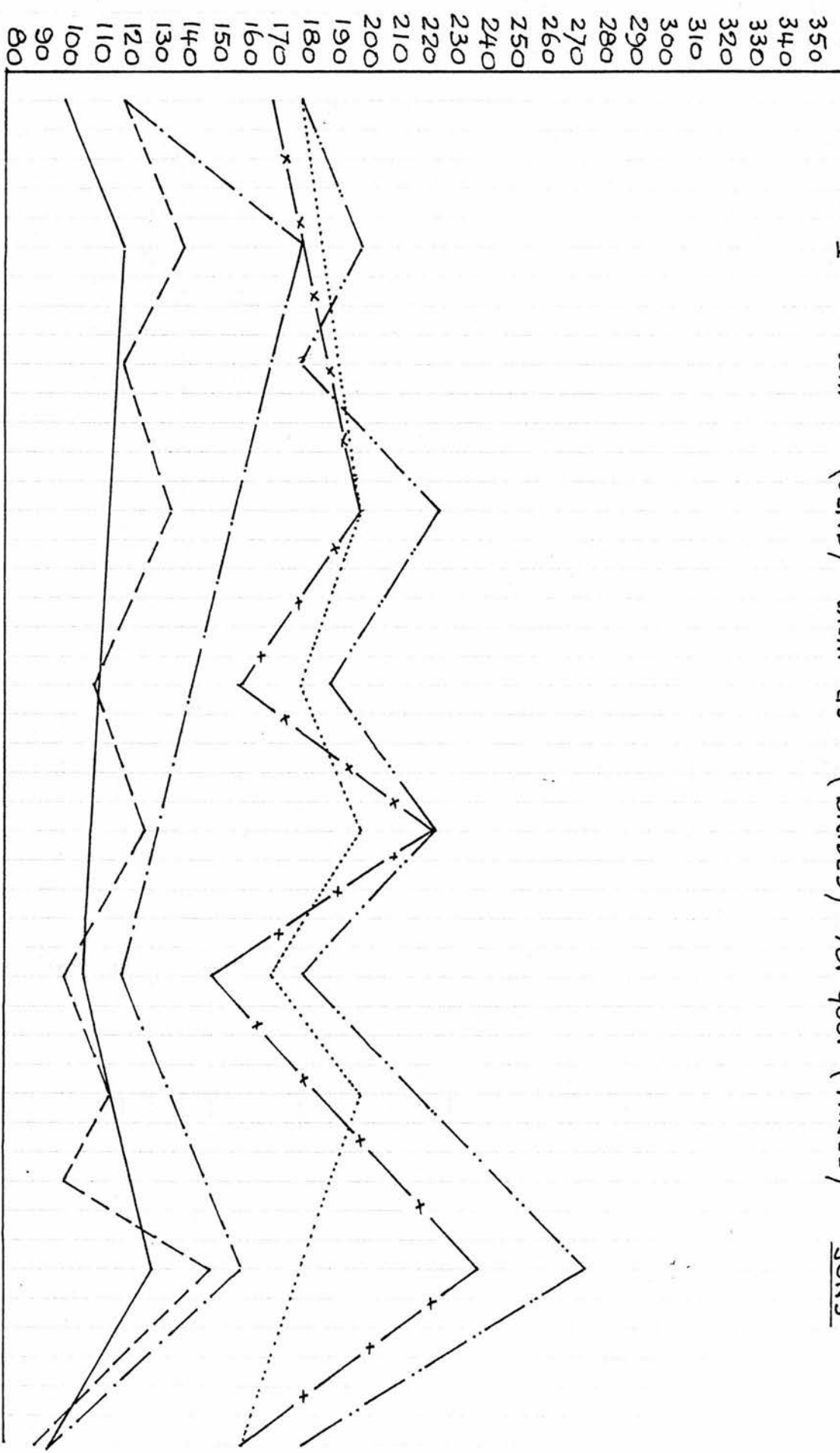
if

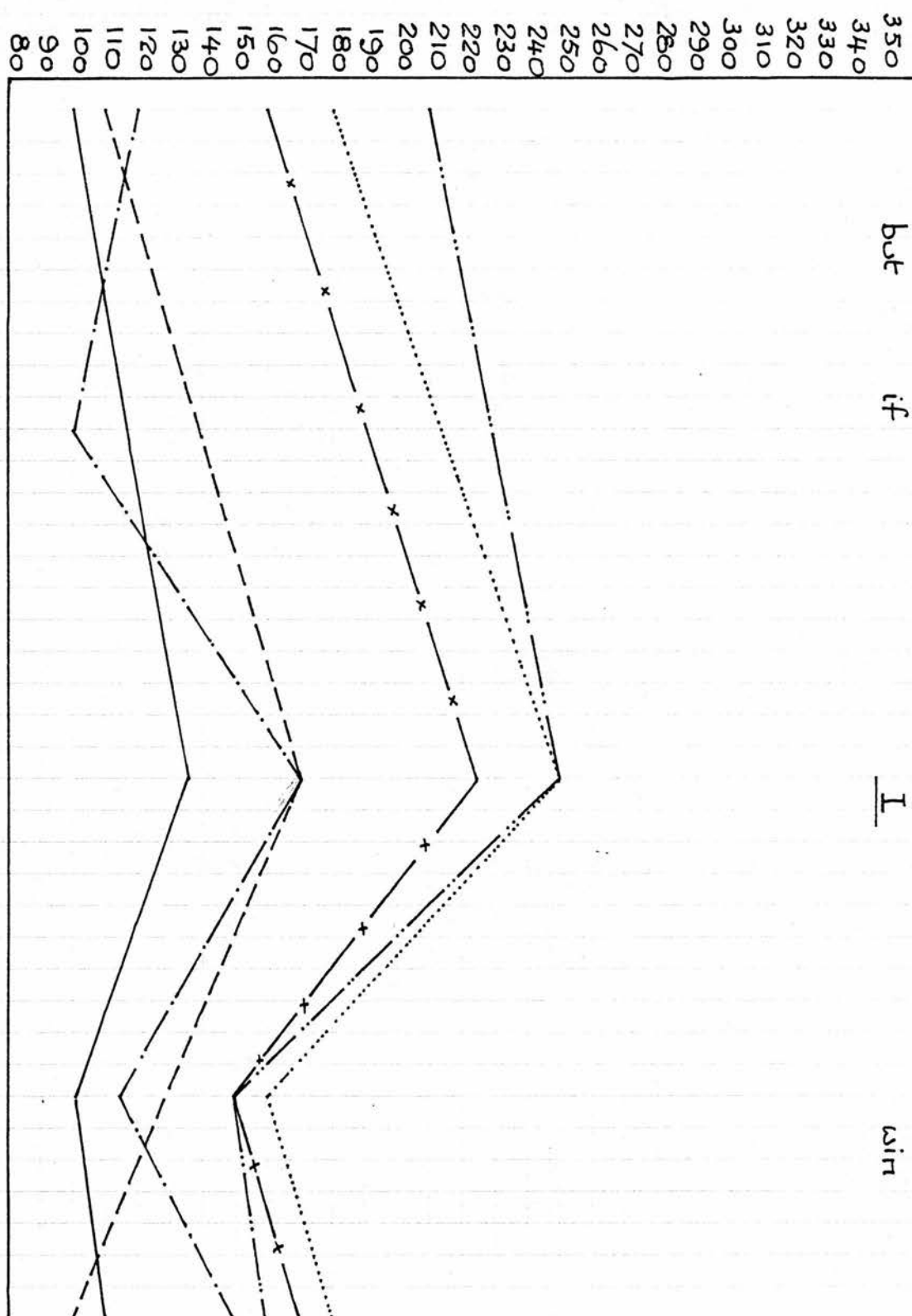
You

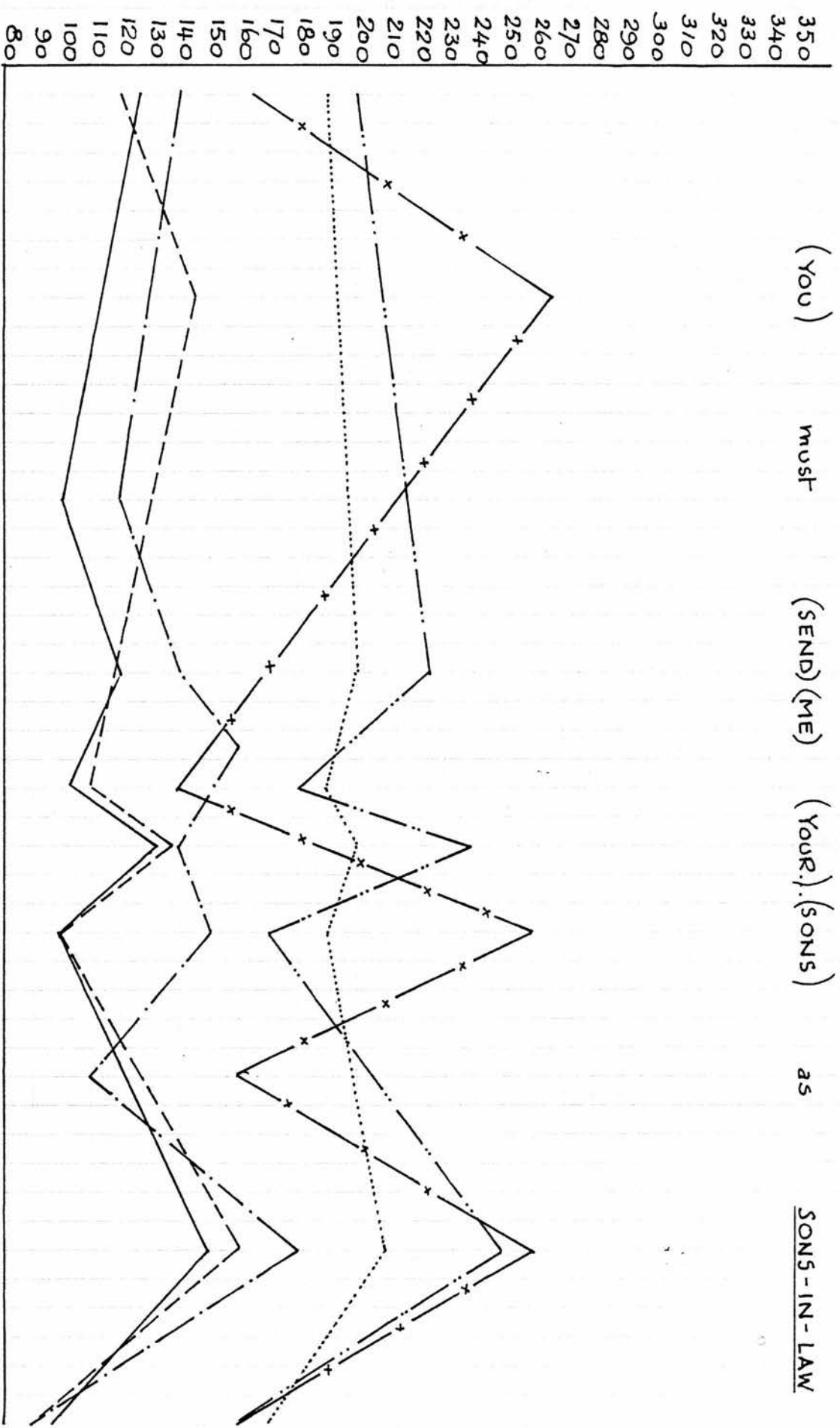
win



I will (SEND) them as (BRIDES) for your (THREE) SONS



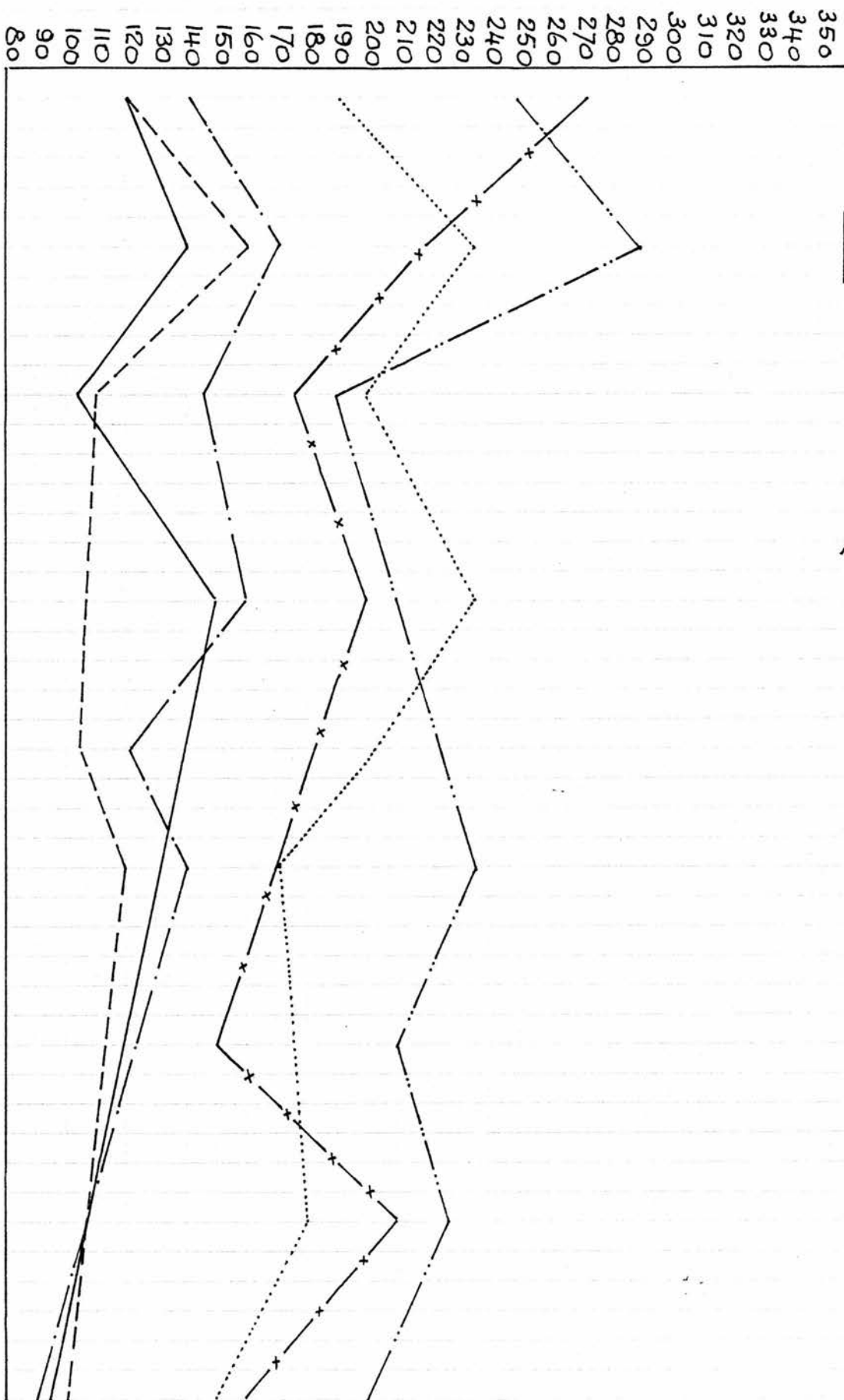




the OLD man

(STROKED) his (BEARD)

for a (WHILE)

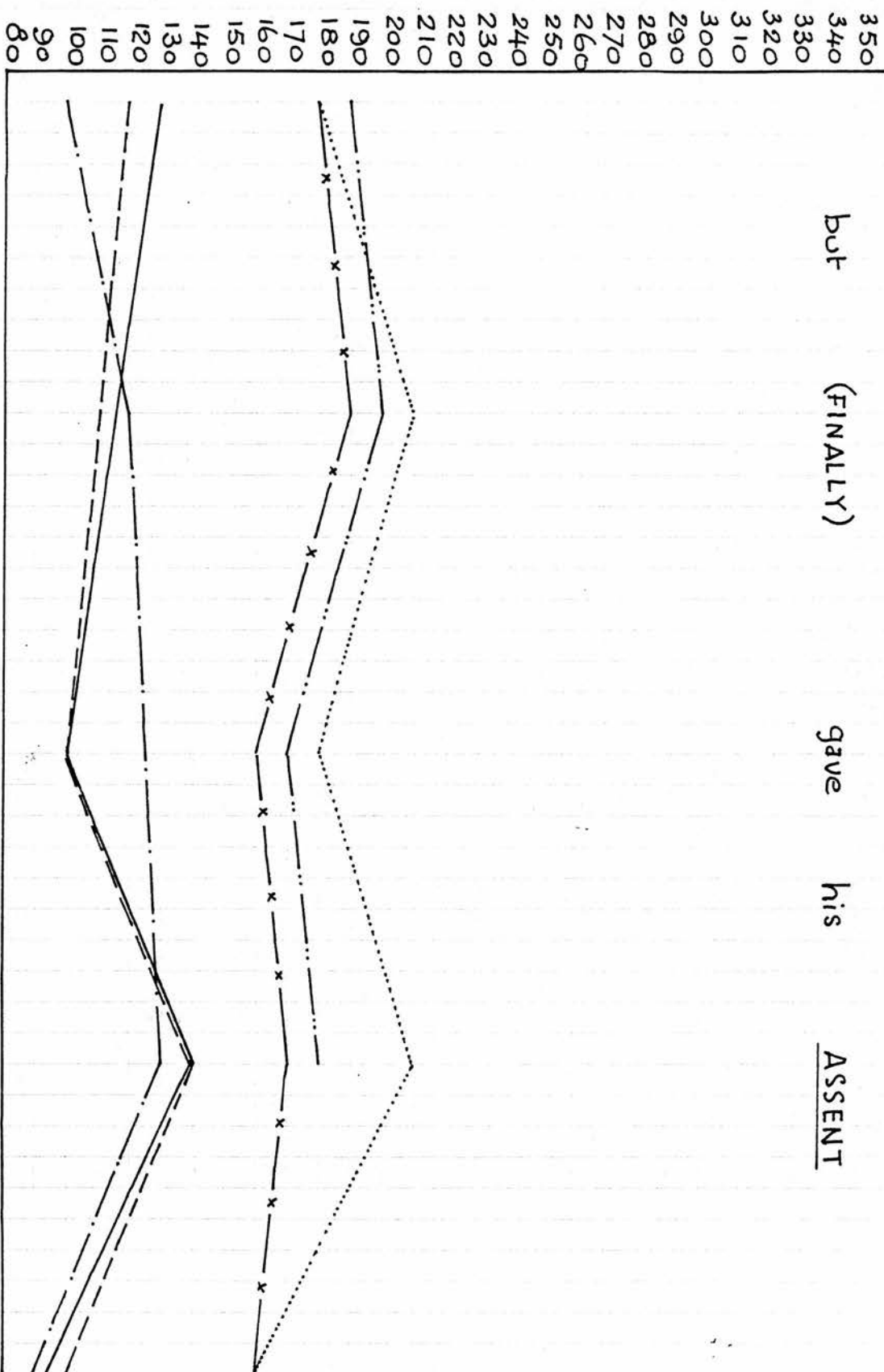


but

(FINALLY)

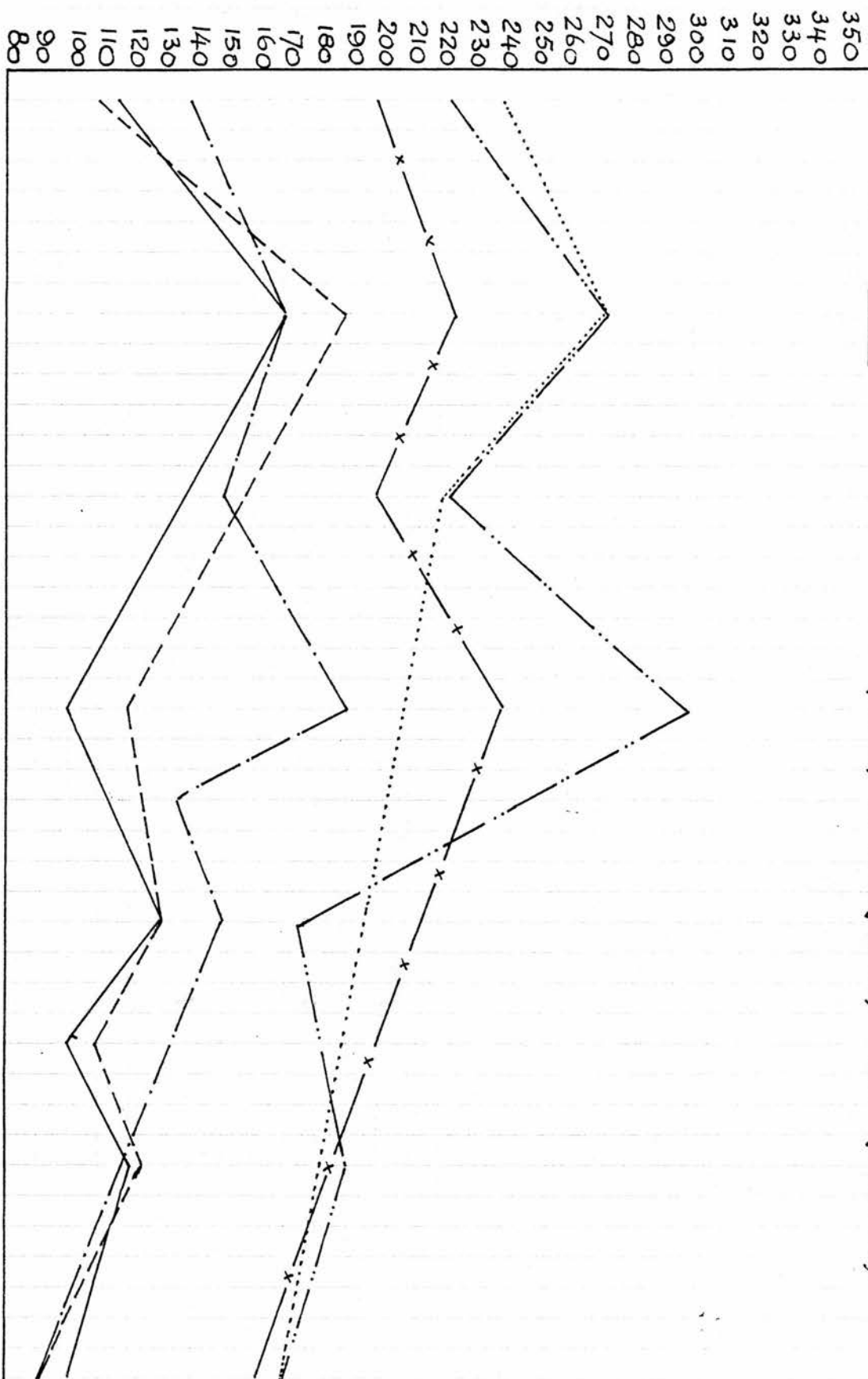
gave

his

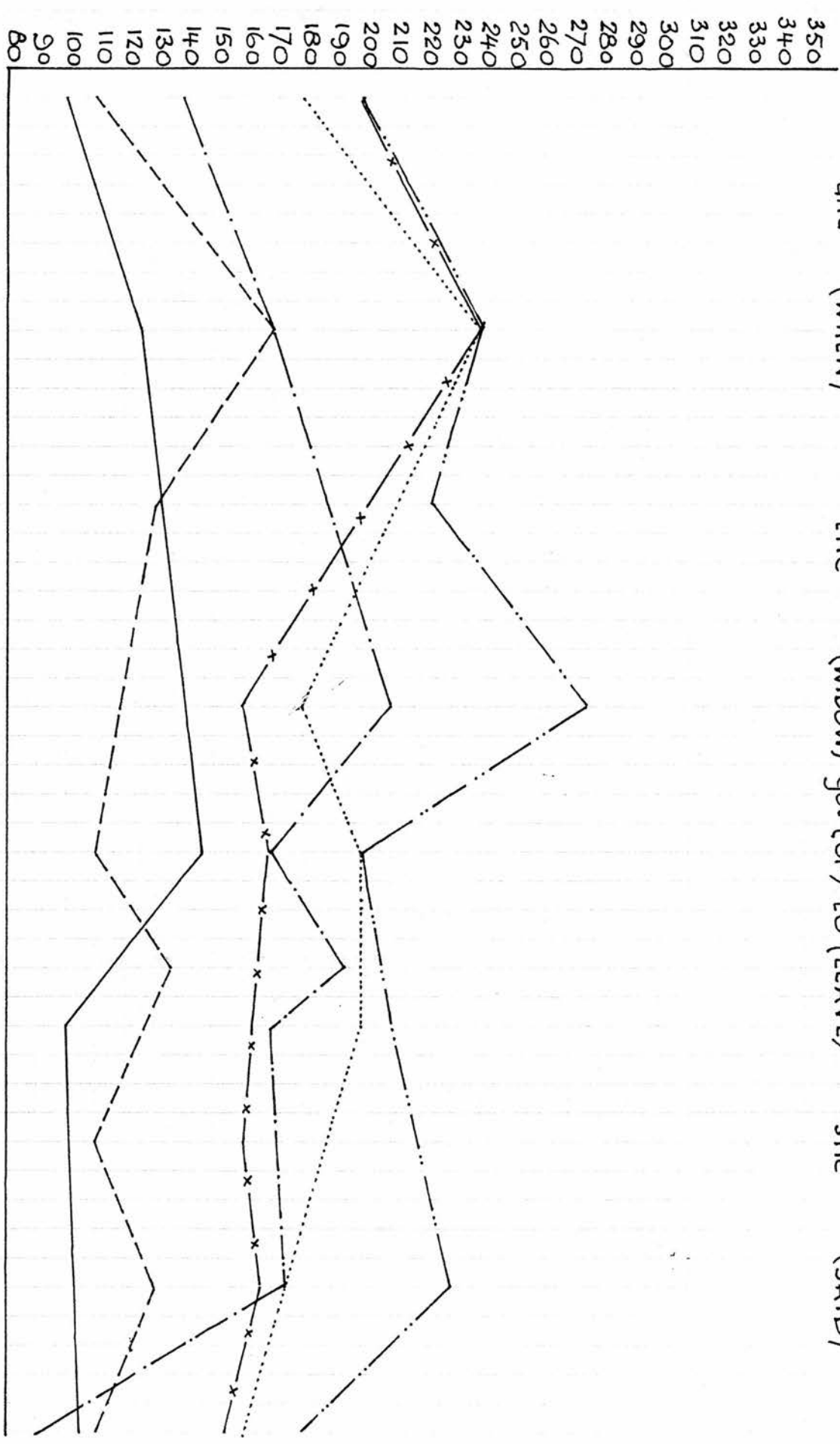
ASSENT

he LOST

(EACH) of the (GAMES) he (PLAYED)

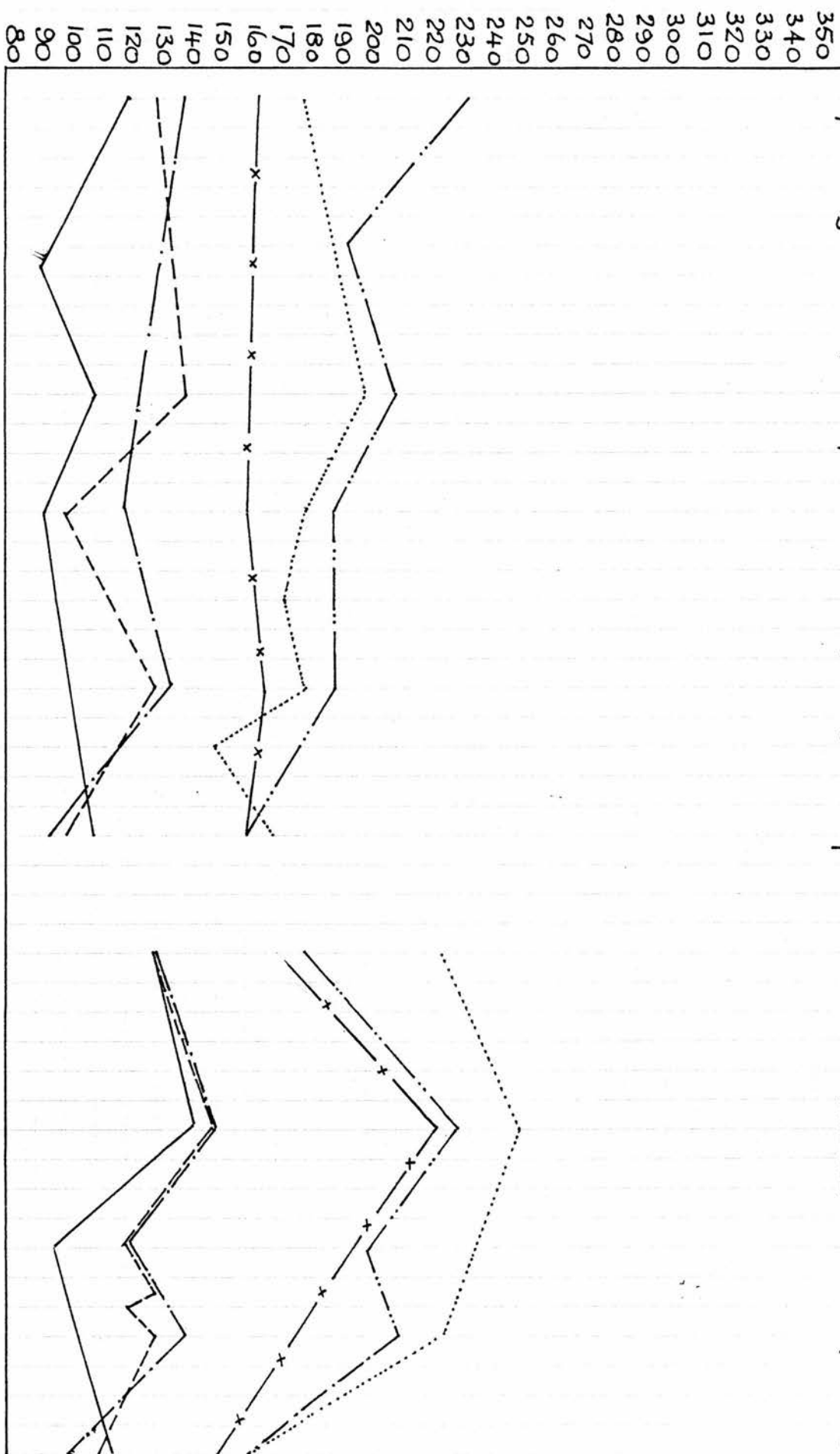


and (WHEN) the (WIDOW) got (UP) to (LEAVE) she (SAID)



pointing (Down) into (Dark) valley

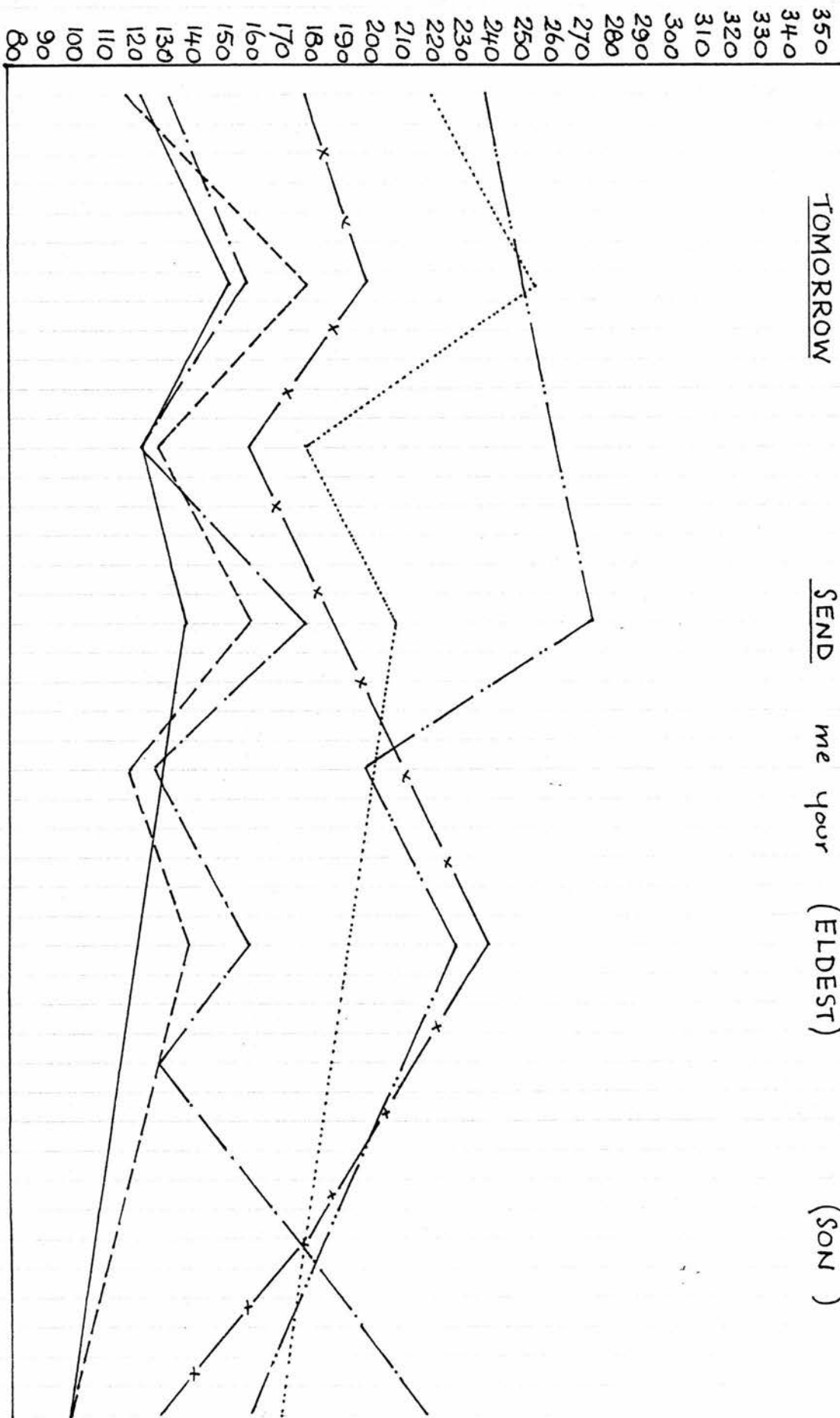
THERE is (MY) (HOUSE)



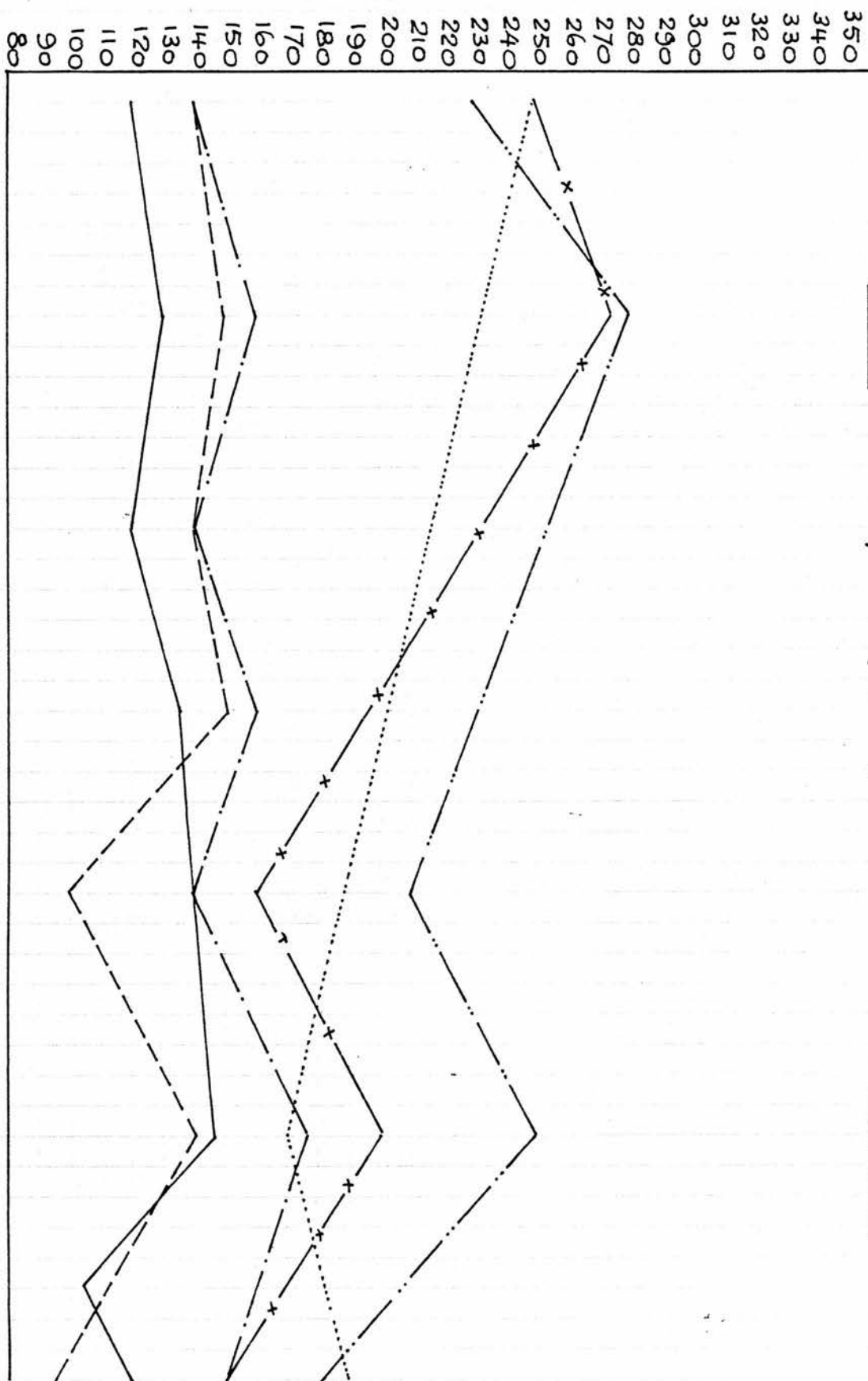
TOMORROW

SEND me your (ELDEST)

(SON)

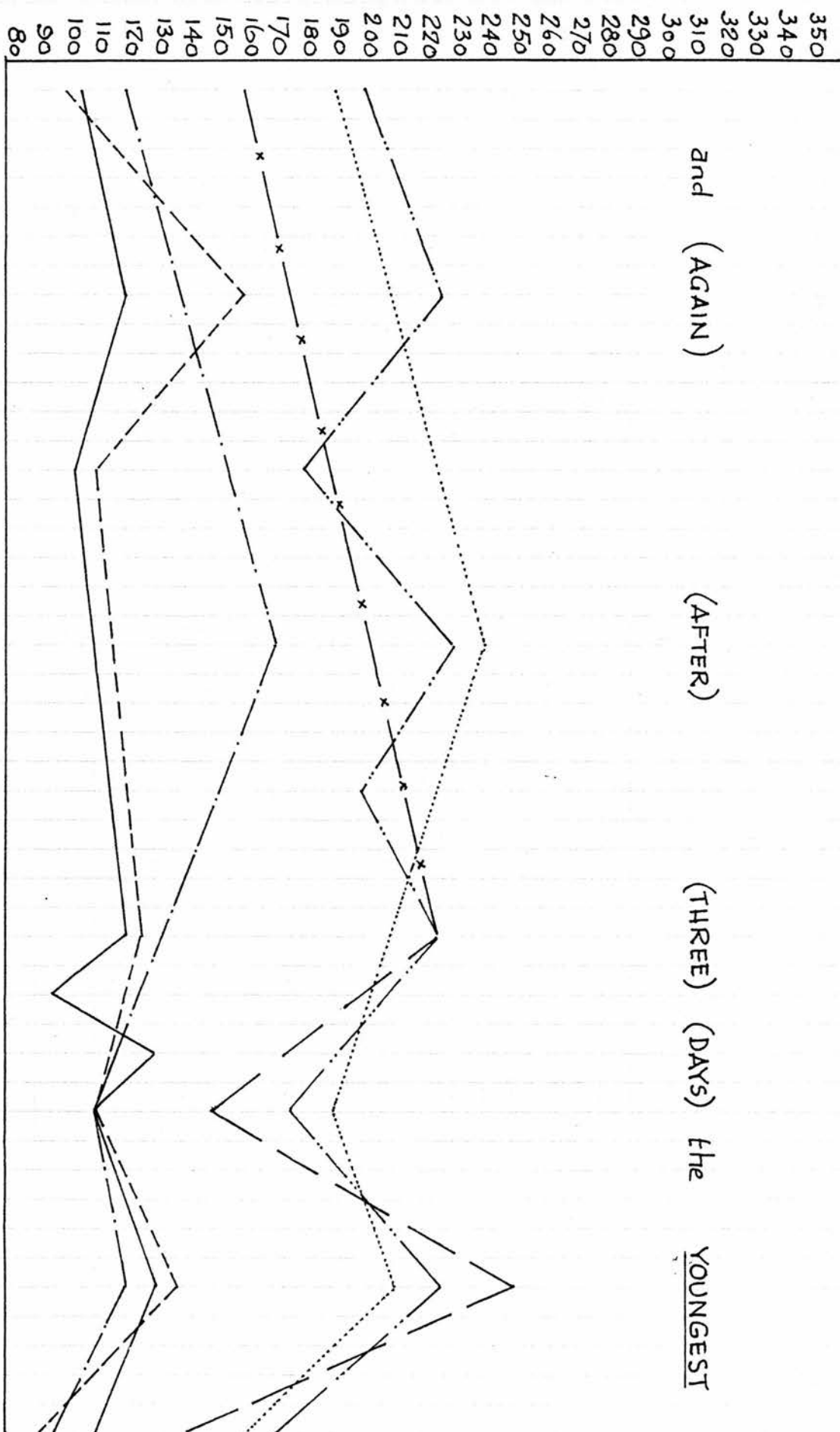


THREE days (LATER) the SECOND



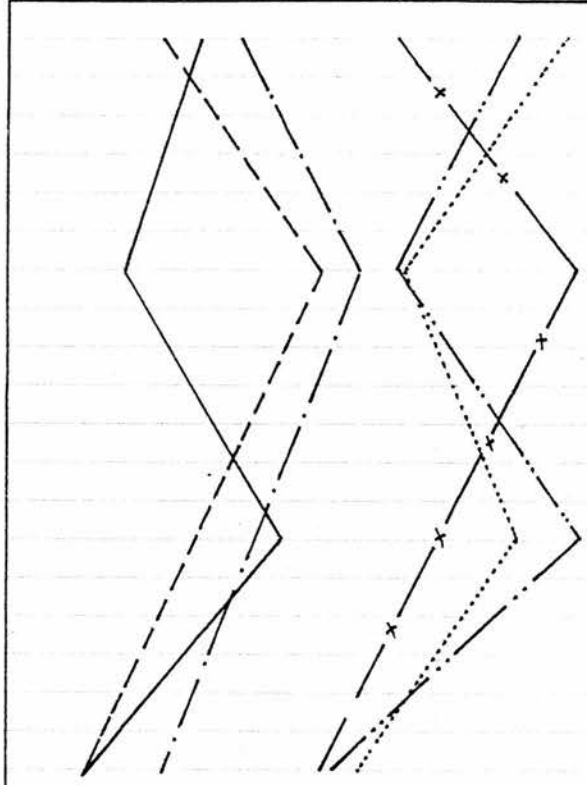
and (AGAIN)

(AFTER)

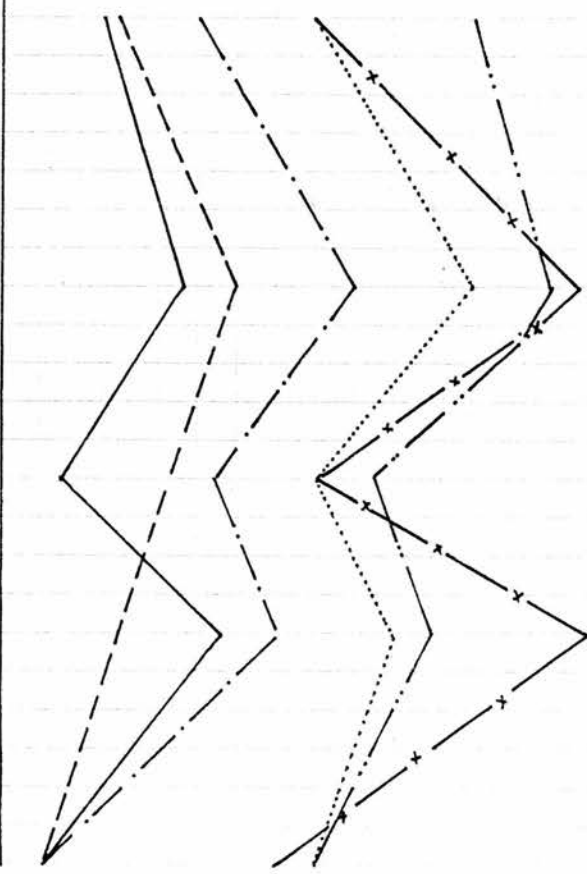
(THREE) (DAYS) the YOUNGEST

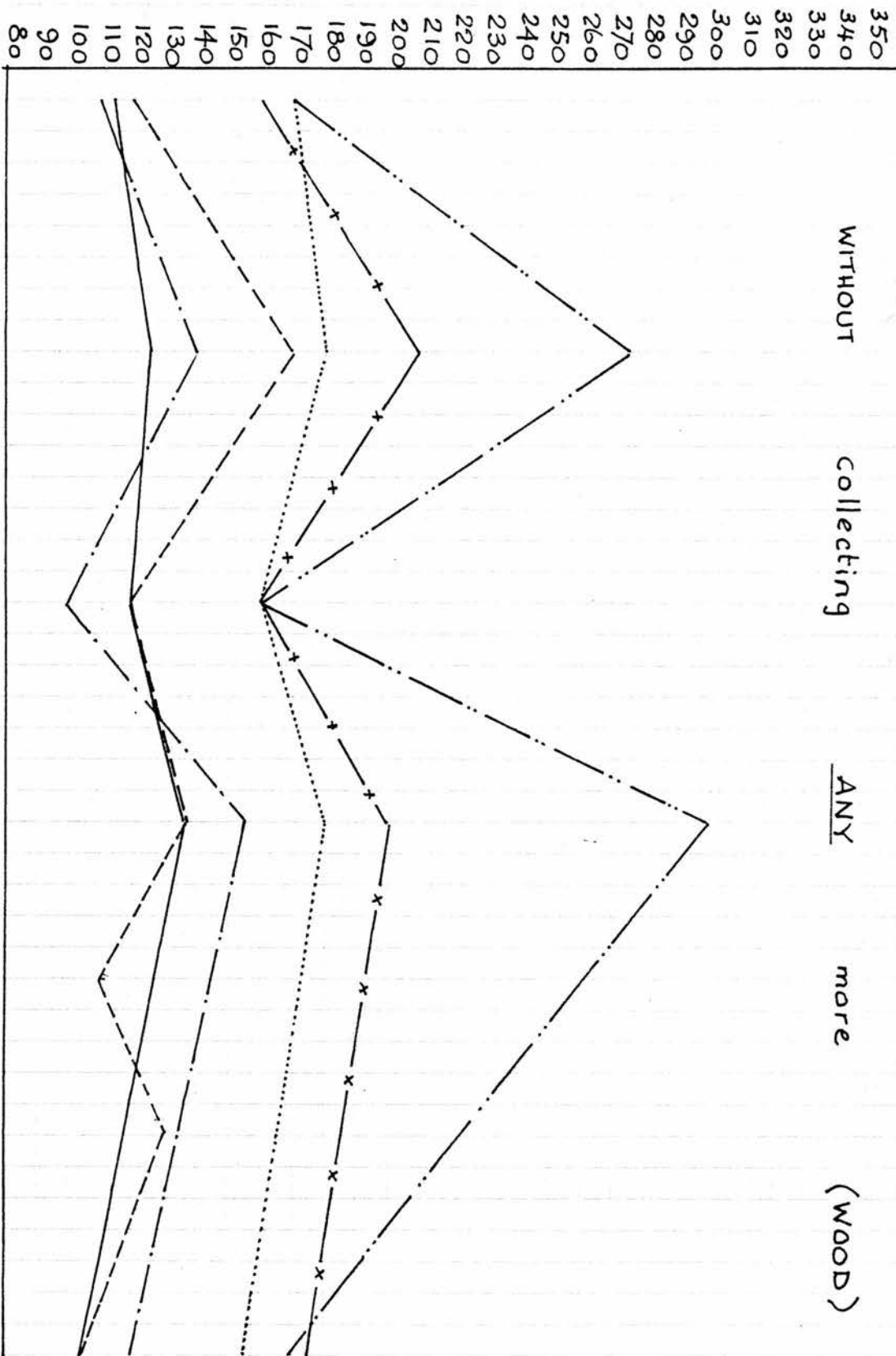
she (THEN) (DEPARTED)

350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80

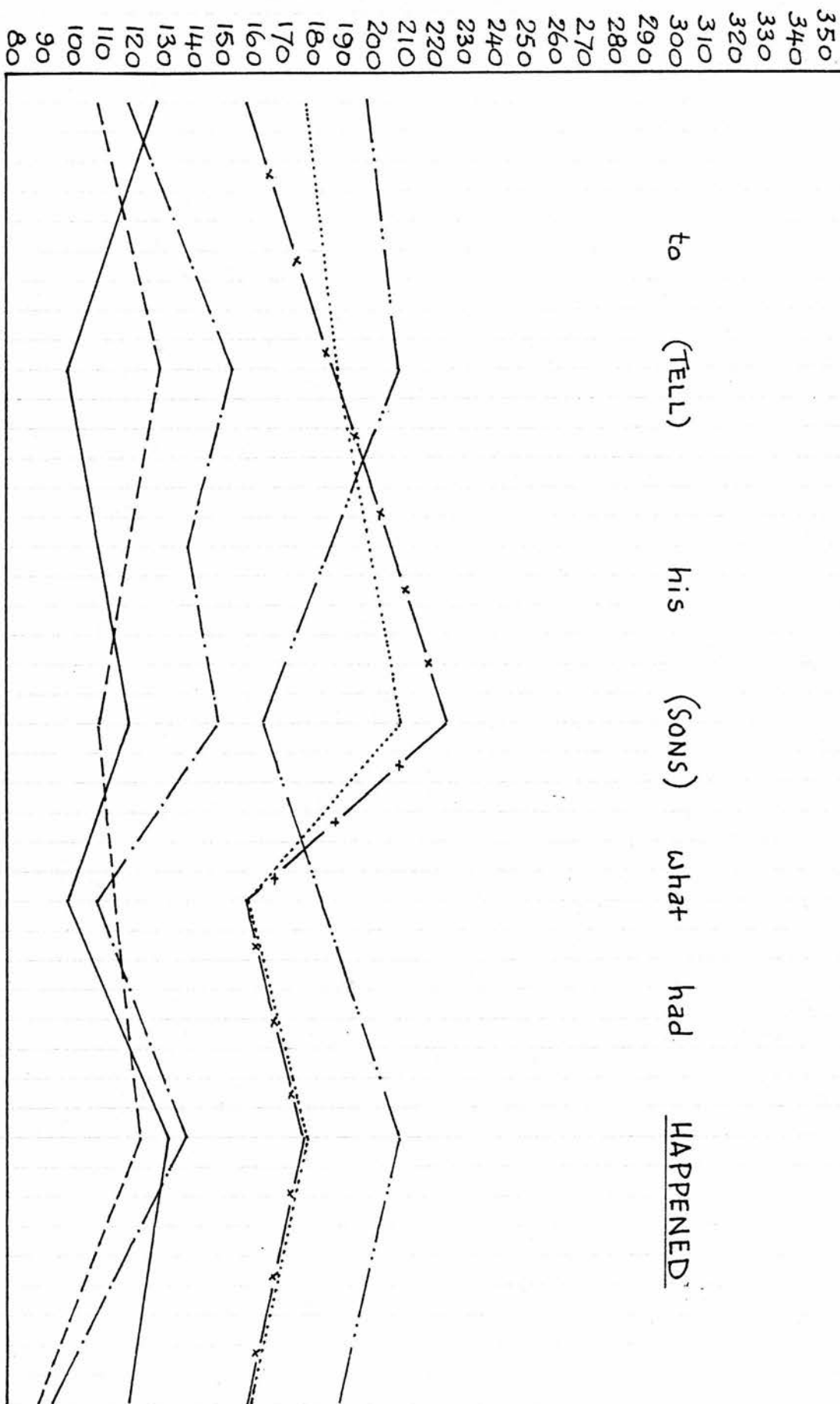


and the OLD man went HOME

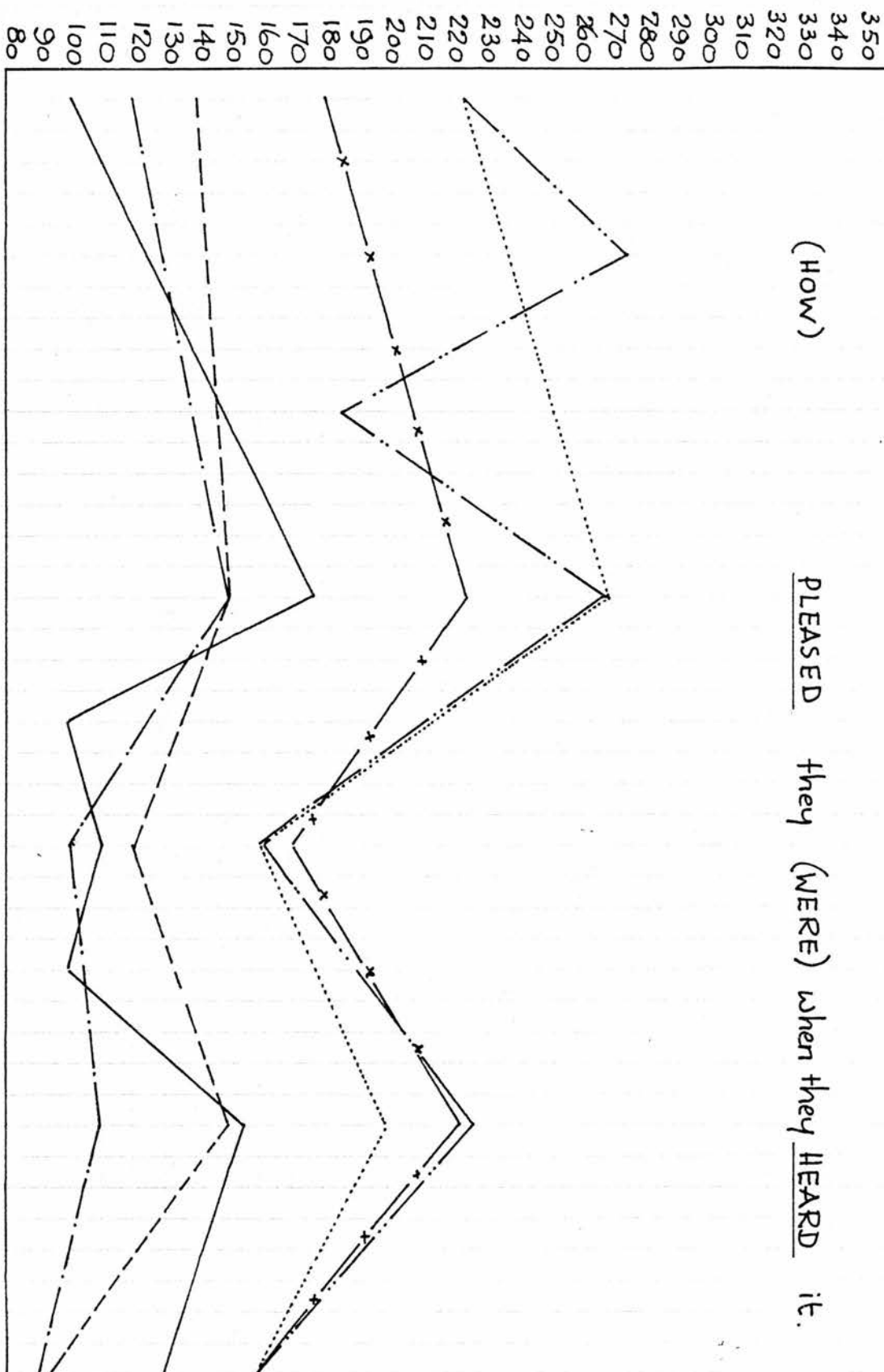




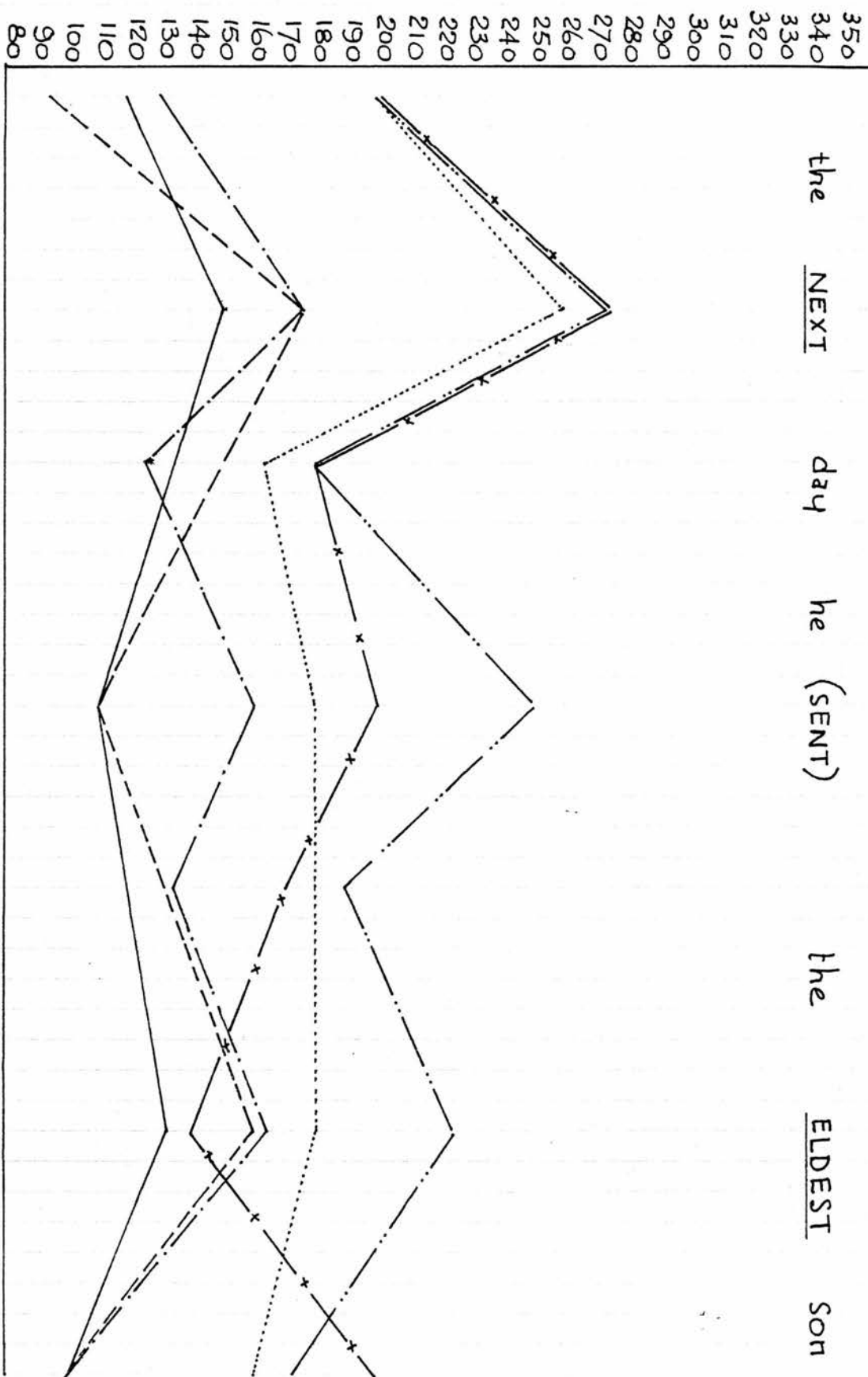
to (TELL) his (SONS) what had HAPPENED



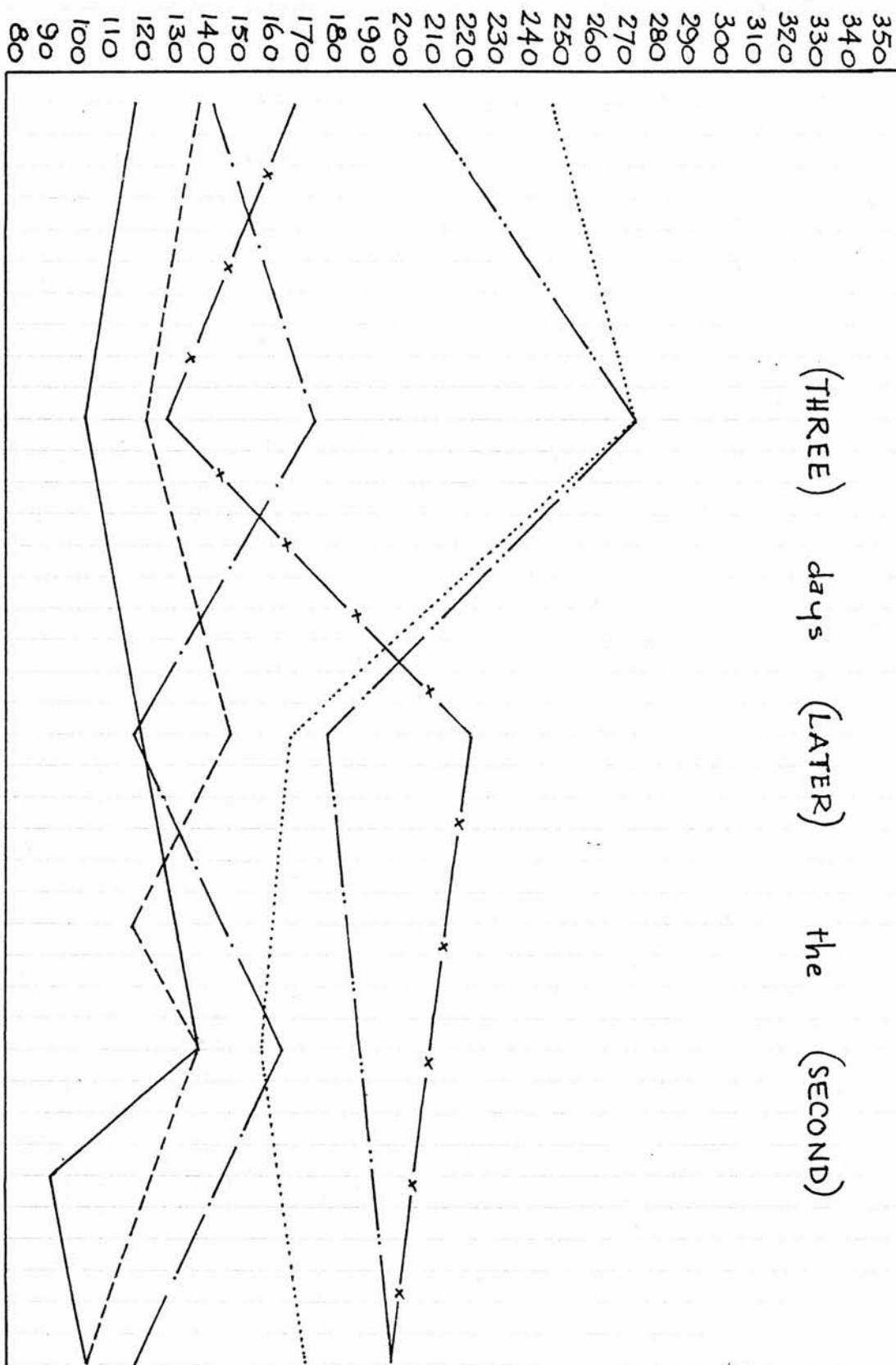
(How)

PLEASED they (WERE) when they HEARD it.

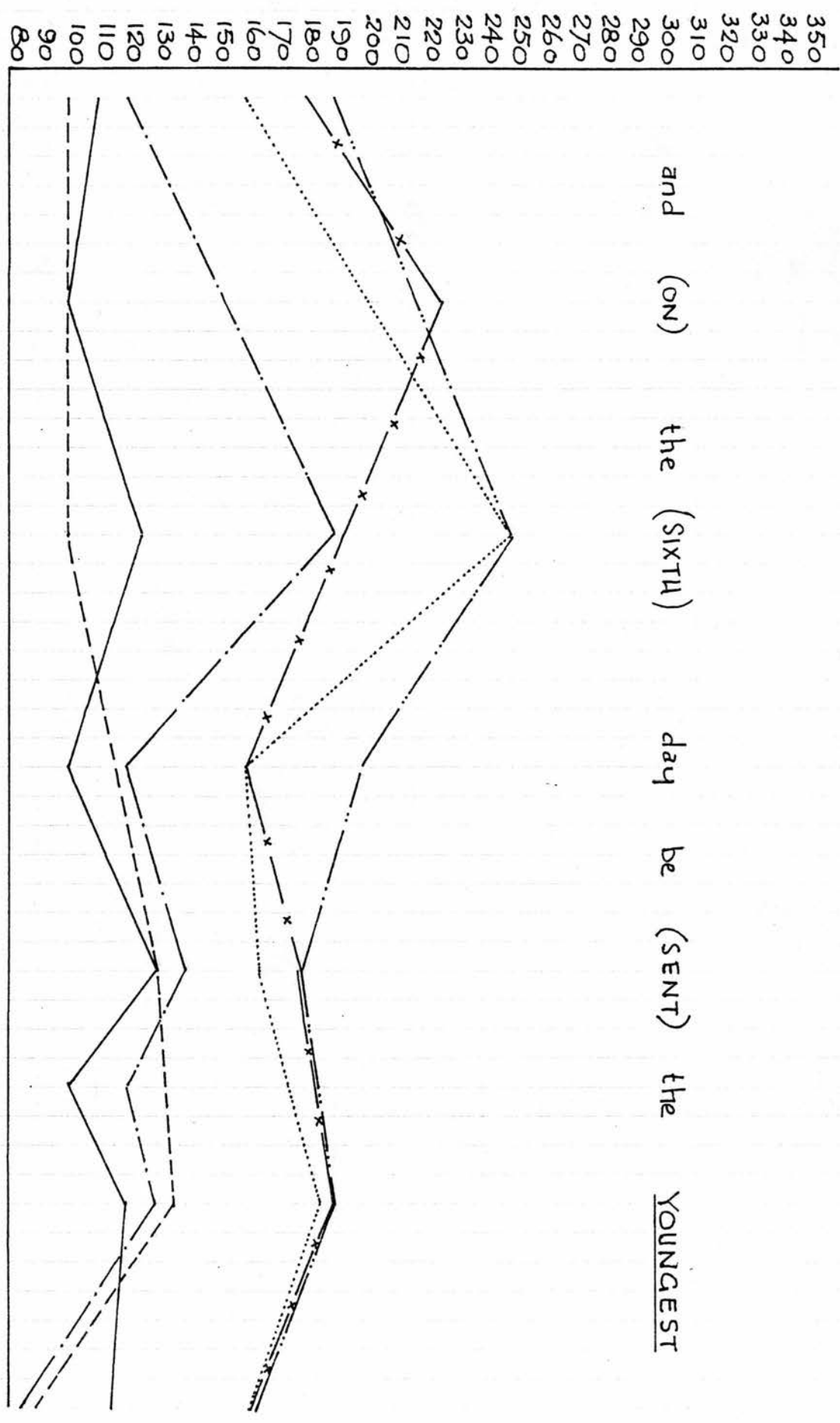
the NEXT day he (SENT) the ELDEST Son



(THREE) days (LATER) the (SECOND)



and (on) the (SIXTH) day be (SENT) the YOUNGEST



Chapter 7.

7.0 An initial examination of contour units in spontaneous speech

In the previous chapter, chapter 6, I examined pause defined units in six read texts. In that chapter I demonstrated that the three contour types contour X, contour Y and contour XY accounted for approximately 85% of the contour units in the readings of these texts. In order to see whether the same contour types occur in spontaneous speech, an initial examination of the spontaneous speech of these same six speakers was carried out. If we subsume the above three contour types under the heading of 'several-peaked' contours, then a numerical comparison can be made between the occurrences of single-peaked contours versus the occurrence of 'several-peaked' contours. This comparison can be set out in table form under the various headings which correspond to the various speech styles examined ; cf. Fig. 1. (the numbers represent the number of pause defined units)

Fig. 1.

	<u>QUESTIONNAIRE</u>		<u>PHOTOGRAPH</u>		<u>CONVERSATION</u>	
	peaks (several)	(single)	PEAKS (several)	(single)	peaks (several)	(single)
11MR	0	0	26	17 (7=1word) (9=1const.) (1=given)	0	0

		<u>QUESTIONNAIRE</u>		<u>PHOTOGRAPH</u>		<u>CONVERSATION</u>	
		peaks (several) (single)		peaks (several) (single)		peaks (several) (single)	
14WF	9	7	14	9	37	8	
		(2=lword)		(5=lword)		(4=lword)	
		(1=listing)		(2=lconst.)		(3=lconst.)	
		(5=lconst.)		(2=given)		(1=given)	
		(3= given)					
16LK	10	11	19	10	27	6	
		(2=lword)		(9=lword)		(3=lword)	
		(1= listing)		(1=lconst.)		(3=lconst.)	
		(5=lconst.)					
		(3=given)					
17AC	12	7	5	10	0	0	
		(3=lword)		(9=lword)			
		(2=lconst.)		(1=lconst.)			
		(2=listing)					
22BS	8	11			172	32	
		(4=lword)				(6=lword)	
		(6=lconst.)				(14=lconst.)	
		(1=given)				(4=given)	
						(4=contrast.)	
33WM	27	9	0	0	30	2	
		(1=lword)				(1=lword)	
		(5=lconst.)				(1=lconst.)	
		(3=given)					

From the above figures it is obvious that, in general, as the situation becomes more informal, the speakers produce longer tone groups which contain more than one peak of prominence. It is interesting to note that all of the single peak contours (apart from 4 in the speech 22BS) occur in specific contexts. These contexts are as follows:

- (a) units consisting of one word only

- (b) units consisting of one constituent (e.g. a noun phrase, proper name, etc.)
- (c) units occurring in a short listing sequence.
- (d) short units which repeat a previous unit such that all constituents are 'given'
- (e) short units which contain a contrasted element.

There are a total of 128 pause defined units containing a single peak, and 124 of those units occur in the above contexts. There are 339 pause defined units which contain at least two peaks of prominence, therefore the contour units which contain several peaks are in the majority. If we consider the fact that the occurrence of almost all of the single peak units can be predicted according to the structure of the units which are realised by single peak contours, then we might suggest that the single peak contour unit is a 'marked' form of contour, and that the unmarked form is a contour which contains several peaks of prominence. The latter are clearly in the majority. These findings support the findings in the read texts. We will now examine a section of spontaneous speech in detail in order to determine whether the contour types described in the text occur in spontaneous speech.

7.1 Contour units in spontaneous speech: Edinburgh English

I propose to examine a stretch of spontaneous speech taken from speaker 22BS. This stretch of speech is taken from the speaker who produced the longest stretch of spontaneous speech (of the six speakers mentioned in chapter 6 and in section 7.0) and who therefore might be considered to be the most relaxed of the six informants. This particular section of conversation is taken from a 'descriptive-narrative' section where the speaker is discussing the South Side of Edinburgh where he has lived all of his life (cf. Appendix 1).

This stretch of speech was measured acoustically in the manner described in chapter 6. Again, I propose to concern myself principally with pause measurements which will define the contour units, and the fundamental frequency characteristics of the contour units themselves.

7.1.1 Pause measurement

Short pauses are represented by +, and if they measure more than 0.3secs, the measurement is given following the pause marker, long pauses are represented by ++, with the measurement following. Units defined by measured pauses are given separate lines of text and numbered consecutively down the left hand side. The

high and low extremes of Fo measurements are given under the appropriate words of the text. Dashes join the extremes which occur within one word. Let us examine the first five units:

1. I regret + putting the people out of the out of
155 200 155 115 190 160 120 115 120 105

the South Side and central Edinburgh you
105 120-110 120-105 110 120-115 120-100 105

know + (0.86sec)
140-100
2. I don't think ++ (1.8secs)
145 175 100
3. especially after the war you know after
165-115 140-110 110 150 110 120 150

the + (0.64secs)
100
4. war when they started the + (0.32secs)
120 120 100 130-100 110
5. redevelopment and the ++ (1.00secs)
105-120-100-120 110 110

In the first section there are five units surrounded by pauses of differing lengths. Two of the pauses (following units 1 and 3) are of the same length as the boundary pauses which demarcated contour units in the read texts i.e. they are between the values 0.6-0.8secs long. There are also two longer pauses occurring after units 2 and 5. After unit 2 the speaker introduces the

notion of the war into the conversation. This element has not been mentioned before and can thus be regarded as a new topic which is being introduced into the conversation. The pause following unit 2 is thus a realisation of a long pause preceding a new topic. The pause following unit 5 can also be treated as a realisation of a long pause since it precedes the introduction of another new topic, the authority.

The last pause in this section which has not yet been mentioned occurs after unit 4 and is 0.32 secs long. This pause is shorter than the normal boundary pause and occurs between a determiner and its head, but then so do the pauses following units 3 and 5. The pause following unit 3 has been referred to as boundary pause, and the pause following unit 5 has been referred to as a long pause. Can we differentiate between the environments of these three pauses?

Unit 3 already contains the element the war, thus the boundary pause occurs during a repetition of this element. In unit 5 the pause again occurs after a determiner, but the head is not realised after the pause. The speaker seems to decide that the authority constitutes his next topic, and he leaves the previous unit incomplete and starts a new topic after a long pause. In unit 4, the short pause occurs between a determiner and its head. The head has not been mentioned previously therefore this unit differs from unit 3, and the head is realised after

the pause, thus differing from unit 5. This short pause, then, looks very like a planning pause described in chapter 4 occurring in the middle of a constituent which is then completed following the pause. Because there is a series of pauses which occur apparently in the middle of constituents (following units 3-5), this series may well be regarded as an example of the 'speaker-maintaining' pause described in chapter 4, where a series of pauses occurs in the middle of consecutive tone groups signalling that the speaker wishes to maintain his role as speaker. But since the pauses all differ in length, and since each unit preceding each pause is different in structure, perhaps they should not be treated as examples of the speaker-maintaining pause. There is a confusion of pause function here, which was not found in the read texts.

We therefore have examples of the pauses which were found in the read texts, but we also have examples of the types of pauses which occur in spontaneous speech (described in chapter 4) which were not found in the text. The pauses discussed so far in the first five units of this section of spontaneous speech are as follows:

A) boundary pause: this pause demarcates the stream of speech into phonological units. These units are the

contour units which were discussed in chapter 6 with reference to the readings of the text. The boundary pause functions as the basic 'chunking' device in speech, dividing speech into units which are easily processed both by the speaker and the hearer. The boundary pause occurs at fairly regular intervals in the stream of speech and will normally be in the region of 0.6-0.8 secs long.

B) long pause: this pause 'chunks' speech into units larger than the contour unit. The long pause will occur preceding the introduction of a new topic into the conversation. Stretches of speech demarcated by the long pause will normally contain several boundary pauses. The long pause will normally be between 1.00-2.00 secs long.

C) planning pause: this pause occurs frequently in spontaneous speech but is rarely found in texts read aloud. The planning pause, therefore, is not regarded as part of the pause system which is operational in speech. It is an incidental pause which occurs when the speaker cannot remember what he was going to say next, or when he has not yet organised the remainder of his utterance into a processible form. The planning pause will normally occur when the speaker has to plan ahead in order to complete a tone unit or tone group. It will typically occur in the middle of a constituent, which

will be completed after the pause. The planning pause will normally be shorter than the boundary pause.

D) speaker-maintaining pause: this pause occurs in the middle of several tone units or tone groups. Its function is to allow the speaker to demarcate units of tone group length without losing his role as speaker. This type of pause can be seen as a variety of boundary pause, which is used in an interactive situation and which is not found normally in readings of a text.

7.1.2 Fo measurement

Having examined the types of pause which can be found in spontaneous speech let us now examine the contour units which are demarcated by the various measured pauses.

There is only one unit out of these first five units which contains a single peak (i.e. unit 2) and this unit is incomplete, therefore what we found to be true in the read texts is also true in spontaneous speech, i.e. that it is more common to have a contour unit containing several peaks of prominence than to have a contour unit containing only one peak of prominence.

Unit 1 starts off with two very high peaks around 200cps., and these peaks are followed by six peaks which are much lower according to the Fo measurements (i.e.

about 120cps.). This contour seems to be very similar to contour X where an initial peak (or in this case two peaks) is greater than any subsequent peaks. In the reading of the text, this contour functioned as the initial contour of a paratone. In this case the contour is the first unit of this section of spontaneous speech where the present speaker (22BS) has just started to speak after the interviewer has finished speaking. Unit 1, then, can also be said to be the initial unit of a paratone.

Unit 2 has been mentioned previously. This is the only unit which has a single peak of prominence. The unit is incomplete and consists of a single clause containing one negative. The speaker focusses on the negative element by peaking on the word don't.

Unit 3 has four peaks of prominence all occurring between the values of 140-160 cps. The initial peak is slightly greater than the following peaks, but the differential is not so great as in unit one. Nevertheless the speaker is introducing a new element into the conversation, i.e. the time element associated with the war, therefore we do seem to have a new cognitive unit which is being realised as a paratone (see chapter 6 for further discussion). But since the initial peak in unit 3 is not as high as the initial peak in unit 1, and since the differential between the initial peak and the subsequent peaks is not so great in unit 3 as in unit 1, we

might propose that there are two different types of cognitive units. Let us propose that when the overall theme of a discussion is maintained, the theme will be realised as a major paratone, and when new elements or topics are introduced within a main theme, then the new topic will be realised as forming a minor paratone. Both of these types of paratone will be delimited by long pauses, but the initial peak(s) of a major paratone will be higher than the initial peak(s) of a minor paratone. Thus if the new topic can be linked to the preceding theme either syntactically or semantically, the new topic will be treated as part of the overall theme and will be realised as a minor paratone with smaller initial peaks than will be found in a major paratone. Unit 1 is therefore the initial unit of a major paratone, and unit 3 is the initial unit of a minor paratone. The overall theme of the discussion is maintained in unit 3, i.e. the speaker continues to discuss rehousing, and unit 3 begins with the element especially which refers back to the preceding units. Unit 3 is therefore 'linked' to the overall theme initiated in unit 1.

Unit 4 contains two small peaks both of which measure between 120-130 cps. This unit conforms to the notion of a tone group which contains two peaks, but it is followed by a planning pause and not a boundary pause.

Unit 5 also contains two small peaks measuring 120

cps each, despite the fact that there is only one open class substantive in this unit.

Units 4 and 5 are examples of the contour XY where the contour contains two or more equal peaks. They occur towards the end of a paratone. The overall structure of these five units can be said to be typical of a paratonic pattern.

The typical structure of a paratone as exemplified by units 1-5 can be stated as follows. The initial unit of a paratone will be a contour of the type contour X, such that the initial peak(s) will be greater than any subsequent peaks. Following the initial unit of the paratone there will be several units of the type contour XY, where each unit contains several equal peaks of prominence. The peaks will typically decrease in height from unit to unit towards the end of the paratone. Within a major paratone there may be several minor paratones. The initial peak of a minor paratone will not be as high as the initial peak of a major paratone. The final unit of the paratone will contain the smallest peaks of the paratone. There will then be a long pause which will precede the beginning of a new paratone. There are no examples of contour Y in this section.

If we examine the rest of the extract in the light of the preceding statements we find that the statements are generally true although the pattern is not always as clear as described in the first section which we have

already discussed. There are several peaks standing out from the surrounding peaks in the next section, but not all of them can be said to signal the beginning of a new paratone.

Each unit will be discussed separately.

6. well the authority more or less made it
 120 135 165-100 120 100 100 140 100
- that everybody was to go outside you know +
 100 125-100 100 100 100 125-95 100 130-100
- (0.68 secs)

This unit which immediately follows the first five units contains a contour of the type contour X, with an initial peak of 165 cps. This peak is considerably higher than the peaks in the preceding unit, but is not as high as the initial peak in unit 1. If we examine the sense of the unit we find that the main theme of the previous unit is being continued with the introduction of a new element the authority. The authority in unit 6, is in fact being substituted for the pronoun they in unit 4. For the above reasons, unit 6 should be regarded as the initial unit of a minor paratone rather than the initial unit of a major paratone.

7. the gardens and houses but + (0.3 secs)
 90 130-110 100 120-90 120

Unit 7 contains two equal peaks of prominence measuring between 120-140 cps. This unit is therefore an example of the contour XY. Unit 7 is very similar to unit 4 since both seem to be followed by a planning pause instead of a boundary pause. As suggested previously in connection with unit 4, these pauses may be examples of 'speaker-maintaining' pauses since there are similar pauses occurring consecutively. This speaker pauses fairly frequently either in the middle of a constituent (cf. unit 4) or between two constituents which are very closely related (cf. unit 7). This type of pausing seems to be a consistent strategy adopted by 22BS which recurs frequently throughout this extract.

8. I would reckon that eighty per cent of
 160 140 120-100 100 140 140 110 120
 the people + (0.78 secs)
 100 120-90

Unit 8 is another contour of the type contour X. The initial peak occurs on I and measures 160 cps, and there are three subsequent peaks measuring between 120-140 cps. The speaker continues the main theme of the major paratone but introduces his own opinion into the discussion. This unit is therefore the initial unit of a minor paratone.

9.	didnae	want	to	go	out	of	the	town
	150-120	160	100	100	100	100	100	125
	didnae	want	their	gardens	they	were	quite	
	100	120-95	100	115-110	110	110	130	
	happy	where	they	were	if	they'd	built	
	130	100	110	120-100	125	115	125	
	houses	in	the	town	(mhm) +	(0.82 secs)		
	100	150	120	105				

Unit 9 is a very long unit with two large initial peaks between 150-160 cps followed by several smaller peaks, and ending with another large peak on 150 cps. This type of contour was treated in the readings of the texts as an equal-peaked contour where the equal peaks were separated by one or two smaller peaks called pivot peaks. In this case, however, there are six intervening peaks separating the equal peaks by a considerable distance. The initial peaks stress the fact that the inhabitants of this area had no wish to leave the area, and the speaker feels very strongly about this fact. His emotions are therefore involved here, and although we could treat this unit perhaps as the initial unit of some kind of paratone I think that in this case it is more relevant to say that the emotional content of this unit gives an extra boost to these initial items. The last peak in this unit occurs on the item in. This element is in fact contrasted since the speaker is stressing the fact that houses might have been built inside the town as opposed to outside the town. The final peak

is contrasted, and for that reason is boosted. Unit 9, then, does contain peaks which stand out from the rest of the peaks within the unit, but these peaks have additional height either because of the emotional content involved or because the item is contrasted. The length of this unit is obviously far greater than any previous tone group, yet the unit is delimited by a boundary pause. The unit could be divided into units of a more typical tone group length (i.e. containing between 2 and 3 peaks of prominence as suggested in chapter 6) as follows (with the peaks underlined):

9. didnae want to go out of the town / didnae want
their gardens / they were quite happy where they
were / if they'd built houses in the town /

In spontaneous speech, then, boundary pauses do not consistently mark tone groups. In unit 9 several tone groups appear to be 'run together' with no pause separating them at all. We might refer to this type of unit as a sense group. The sense group will contain several tone groups and will be delimited by boundary pauses. The domain of the sense group will therefore be greater than the domain of the tone group.

10. which they're doing now after +
160-125 160-145 160-120 165 115-110

(0.34 secs)

11. about eh + (0.86 secs)
 110 105
12. (mhm) you know + (0.38 secs)
 160 120

Unit 10 is fairly high throughout with the unstressed syllables very high in the speaker's pitch range (they're measuring 160 cps). The unit starts with the word which, therefore this unit cannot be said to be the initial unit of a new paratone. A new paratone is correlated with a cognitive unit of some kind, and unit 10 starting on which is obviously closely related to the previous unit. The speaker is still very involved with what he is saying and feels very strongly about this particular aspect of the subject. This emotional involvement has the effect of boosting the peaks of prominence and the unstressed items with the result that the whole contour is realised high in the speaker's pitch range, with peaks on four out of the five words in this unit. Unit 10 is bounded by a planning pause which interrupts a constituent. The speaker tries to complete this constituent during units 11 and 12. He uses the filled pause eh and the lexical filler you know to maintain his role as speaker, and finally completes the constituent in unit 13.

13. thirty years too late so to speak
 145-120 125-115 140 120-150 130 80 110-95
- you know (mhm) + (0.62 secs)
 100 135-110

Unit 13 has four equal peaks between the values 135-150 cps, and is a normal mid-paratonic contour unit, of the type XY. One interesting feature of this unit is the presence of the lexical fillers so to speak and you know within the unit. This seems to indicate that the speaker is approaching the end of this theme since he is showing signs of 'running out' of subject matter. There is also quite a lot of movement on the final lexical filler you know which falls from 135-110 cps. This is an unusual amount of movement for a lexical filler and also seems to function as a finality marker (cf. unit 20).

14. that's what I regret especially central
 165-180 110 110 120-90 155-110 120-115
- south Edinburgh + (0.8 secs)
 115-110 115-130-85
15. em ++ (1.5 secs)
 140-115

There were indications in unit 13 that the speaker might be about to conclude this thematic section, yet unit 14 is not a contour containing the lowest peaks of prominence in the paratone, on the contrary, the first peak is the highest in the paratone apart from the initial unit.

is of the type contour X, with the initial high peak on the deictic element that's. This unit is functioning as a summarising unit. The speaker is referring back to the theme of the major paratone initiated in unit one. The speaker uses the word regret, which formed the initial peak of the first unit, thus 'rounding off' the paratone ^{using the same} L₁ lexis. The final end-point is the lowest point in the paratone indicating finality. There is then a long filled pause (lasting 2.9 secs altogether including the duration of the vocalisation) which precedes the initial unit of a new major paratone. The following units seem to constitute a paratone, therefore they will be examined together as a cognitive unit.

16. I would I reckon with taking the people
 110 125 160 180 120 200 120 120
 out they've + (0.36 secs)
 140-120 125-120

17. they've lost the community + (0.82 secs)
 125 165 110 130-140-90

18. no community spirit at all I don't think
 180 130-90 110 110 120 120 180 140
 you've the same spirit in the new housing
 100 100 130 125 100 100 125 110-90
 schemes + (0.7 secs)
 110-90

19. as you had in the old eh + (0.6 secs)
 100 110 115 90 90 120 100

20. (did eh...) type of dwellings you know ++ (1.8 secs)
 110 100 110-100 100 130-100

The pattern displayed by the last five units is very similar to the pattern displayed by the first five units of this extract.

Unit 16 is a contour of the type contour X with two initial peaks at 180 cps and 200 cps. These initial peaks are then followed by two smaller peaks. Thus unit 16 is typical of the first unit of a new paratone.

Unit 17 contains two peaks both of which fall approximately 50 cps. The first peak falls from 165-110 cps, and the second peak falls from 140-90 cps. Unit 17 therefore seems to be a typical mid-paratonic unit containing a contour of the type XY with peaks between 140-160 cps.

Unit 18 contains two large peaks at 180 cps, both of which occur on negative elements. These negative elements may be regarded as the focal points of this paratone since the speaker is emphasising the lack of community spirit in the new housing schemes. This unit is larger than a normal tone group and might be subdivided into tone groups containing 2 or 3 stressed syllables as follows:

no community spirit at all / I don't think you've
the same spirit / in the new housing schemes / +

Unit 18 is therefore another example of a sense group which contains several tone groups which are not

bounded by pauses.

The last two units, unit 19 and unit 20, follow the paratonic pattern observed in the first five units of this excerpt. Both units contain two peaks between the values 110-120 cps. These peaks are the smallest peaks of the paratone. The phrase you know is realised with the movement 130-100 cps, similar to the amount of movement on the same item when it occurs in unit 13. Movement on this lexical filler therefore seems to function as a finality marker when the lexical filler occurs at the end of a unit. As it occurs in unit 20, we seem to have an example of the contour Y, with the final peak greater than any preceding peaks. This is the first example of contour Y in spontaneous speech, but it occurs on the last unit of the speaker's turn. Thus contour Y as it occurs on unit 20 is also a marker of finality, but not finality of a paratone, rather finality of turn.

Throughout the whole of this extract the baseline, formed by the unstressed syllables, remains very consistent measuring between 90-110 cps. The only variation of the baseline occurs in units 1 and 16 where the baseline is raised and measures 115-120 cps. Units 1 and 16 stand apart from the other units in this extract for the following reasons:

- (a) the baseline is raised

- (b) the initial peaks of these units reach 200 cps forming a contour of the type contour X. The initial peaks constitute the highest peaks in the extract.
- (c) unit 1 is the first unit of the speaker's turn, and unit 16 is preceded by the longest pause in the extract.

The above facts support the notion that units 1 and 16 are the initial units of two major paratones which differ from the initial units of minor paratones. The initial units of minor paratones are also signalled by the presence of a contour of the type contour X, but the peaks are not as high as the peaks of a major paratone, and the baseline is not normally raised. The initial unit of a minor paratone will also be preceded by a long pause.

7.1.3 Summary of findings in Edinburgh spontaneous speech

From a detailed study of this section of spontaneous speech, various features emerge which serve to differentiate between a spontaneous speech style and the reading of a text. The realisation of the pause system in spontaneous speech is not as straightforward as the realisation of the pause system in the readings of the text. Occurrences of planning pauses and speaker-maintaining pauses interfere with the formal pause system as described in

chapter 6. Similarly, if we examine the realisation of paratone structure in spontaneous speech, the structure is not always clear. In the above extract, two clear realisations of paratone structure can be observed occurring between units 1-5, and units 16-20. The middle section of the above extract contains high peaks which function in various ways. They signal new minor paratones, emotional involvement, and contrast. Thus high peaks which occur initially in a tone group are not unifunctional as they were in the texts. They are multi-functional, and, as such, each function may not be unambiguously identified on every occasion, and confusion may result. In the above extract there are also various realisations of paratone final units. There are units which contain the smallest peaks of the paratone which indicate finality; there are units containing high peaks which indicate finality by summarising the topic or theme of the paratone; there are units of the form contour Y which indicate finality of turn; lexical fillers also appear to indicate finality when they occur finally within a unit, especially when they contain some movement. This variety of finality markers was not observed in the readings of the texts.

In spontaneous speech we find one cue functioning in various ways (high initial peaks), and we find several cues functioning ⁱⁿ the same way (marking finality).

Thus spontaneous speech seems to exhibit more complex phenomena than the text readings. The same underlying systems can be seen to be present in both spontaneous speech and text reading, but planning phenomena, turn-taking, interaction or role phenomena interfere with the underlying system in spontaneous speech with the result that there is more likely to be a confusion of cues and / or functions in spontaneous speech than in reading style of speech. More 'parameters' are involved in spontaneous speech compared with a reading style of speech since a reading style requires no planning and involves no competition for turn-taking.

7.2 Contour units in spontaneous speech: Thurso English

In order to compare contour units across different accents, an extract of speech from a speaker of Thurso English was subjected to the same analysis as had been carried on the extract taken from 22BS. This extract is from the speech of a 12-year old boy who is retelling a story which he has just heard. This extract is an example of narrative style with rare interruptions from the interviewer and can thus be compared to the descriptive narrative described in section 7.1. The extract is full of fairly lengthy pauses and false starts, therefore possible contour units will be numbered when they have communicative content. The extract is as follows, with

elements realised on peaks underlined; (the interviewer's comments are given in brackets):

1. it's about this man and a wifie + (0.64secs)
225 400 250 300 225 225 275-225
2. they've always wanted a bairn + (0.88secs)
250 325 300 225 300-180
3. they couldnae get one + (1.12 secs) (mhm)
225 325-225 300 190
4. so + in his dreams this man like + (2.8secs)
250-225 250 230 325-210 260 300 225
(shh if there's any giggling you get put out)
5. he + (0.6secs) + hear- heard in his dreams +
275 300-275 300 225 225 300-200
(0.52secs)
6. if he walks + eh (0.98 secs)+ mm (1.12secs)+
250 250 275 270-250 230
for +(0.96 secs) + at Dunnet Head + (0.86secs)
225 220 310 350-200
7. for + a year + (0.64secs) and one day + (0.62secs)
225 225 275-200 200 240 190
8. if he walks up and down every 'fore
225 225 300 270 250 240 270
sunset + (2.42 secs)
275-200
9. well he'll get a bairn + (1.2 secs)
225 230 250 200 275-180
10. and eh the last day + before he came
225 225 245 270-225 190-240 225 225
hame + (0.48 secs)
245-190
11. he saw a bairn + (1.2secs)+ on the rocks
250 190 245-200 200 200 250-200

12. and eh + (0.83 secs) + $\frac{\text{taken}}{250}$ it $\frac{\text{hame}}{225 \ 250-200}$
 190 190
- and $\frac{\text{hurried}}{135}$ tae his $\frac{\text{wife}}{240-200}$ + (0.7 secs)
 190 190
13. and they both got $\frac{\text{dressed}}{230-200}$ + (1.6 secs)
 200 200 240 190
14. and eh ++ (3.06 secs)
15. they + (0.8 secs)+ $\frac{\text{went}}{300}$ tae the $\frac{\text{church}}{280-230}$ + (0.56)
 275 300 300 250
16. and eh + (1.4 secs)
 225 225

The system operating in the above sample from Thurso looks very similar to the one described for Edinburgh. Pauses define units, and each of these units is a contour containing two peaks of prominence or more. This extract begins with a contour unit of the type contour X, with the initial peak measuring 400 cps (the highest peak of the above text). There are then three units which might also be called contours of the type contour X except that the differential between the peak measurements is much smaller than that found in the first unit. In the first unit the differential is 100 cps, and in the following three units the differential is only 25 cps. These units might therefore be referred to as contour units of the type XY for this speaker who possesses a much wider pitch range than that of 22BS. There are then various units interspersed with planning pauses, but the general pattern of units of the type XY continues

through to unit 13, with the peaks gradually decreasing. Units 11, 12 and 13 contain the smallest peaks of this paratone. There is then a very long pause, which for this speaker lasts 4.66 secs plus a vocalisation. Unit 15 then starts a new paratone indicated by raised peaks and a raised baseline.

The general pattern of the paratone as described for the Edinburgh extract of speech applies very well to this extract of Thurso speech.

7.3 Contour units in spontaneous speech: Glasgow English

We will now examine a short extract of speech from Glasgow. This extract is taken from a conversation between myself and a woman in her forties. The extract is given in full in Appendix 2. If we examine this sample closely we find that there are units which are surrounded by pauses. In the Glasgow sample, however, these pauses are shorter than the boundary pauses in the Edinburgh sample and often seem to be realisations of planning pauses; cf. the following examples:

1. em + my uncle's coming home from Canada +
 (0.5 secs)+ on + (0.46 secs)+ Sunday + (0.54 secs)+
 he's due in + (0.52 secs)
7. he was born in the Highlands and he wants to go back
 +(0.32 secs)+ and see + (0.42 secs)+ the Highlands

while he's at home +(0.42 secs)

The pauses measured vary from 0.18 secs to 1.5 secs in this particular extract. There are 22 pauses altogether, 16 of them are less than 0.54 secs long (13 of these measuring between 0.3-0.54 secs long); 5 pauses measure between 0.68-0.86 secs and there is one pause which measures 1.5 secs. This distribution would seem to correlate with the distribution of pauses in both the Edinburgh and the Thurso samples. The majority of pauses are fairly short boundary pauses, and there is a small group of long pauses presumably marking paratones.

If we examine the extract as a whole we find that the long pauses do correlate with changes in topic. The first two long pauses occur at the end of the following section:

em + my uncle's coming home from Canada + on Sunday+
 he's due in + (how long has he been away, or has he
 just been away?) oh no they lived in Canada eh he
 was married to my mother's sister + well she's been
 dead for a number of years now and he's remarried +
 (0.86 secs) but he's coming home they're coming home
 on Sunday + for three or four weeks + (0.82 secs)
 (+ denotes a pause of less than 0.54 secs)

The topic of this section is given in the first unit of the utterance (my uncle's coming home), and 'rounded off' in the final unit (they're coming home on Sunday for three or four weeks). The first long pause occurs after the speaker has explained the relationship between herself and this uncle from Canada, then the second long pause occurs after the summarising unit (unit 4 in Appendix 2).

The topic following this section is again announced and rounded off by the lexis, but it is not delimited by a long pause, instead the interviewer interrupts with a question as follows:

so we hope to see them while they're here + he wants to go back up to the Highlands + he was born in the Highlands and he wants to go back + and see + the Highlands while he's at home + (whereabouts in the Highlands do you know?)

The topic of this section is announced by the phrase while they're here, then rounded off by the phrase while he's at home. A new topic is introduced by the insertion of the question given at the end of the above section, and this topic involves the remainder of the extract:

now I'm not very sure I thought it was Perth +

(0.78 secs)+ but + the way he spoke + (1.5 secs)+
 he seemed to feel it was + further North + but he
 maybe considers + Perthshire + Doune you know +
 (0.68 secs) + I thought they came from Doune +
 (0.68 secs) (that's Perthshire)

Again the topic is announced in the initial unit of the section (I thought it was Perth) and rounded off in the final unit (I thought they came from Doune). But this section is more complex than the previous sections. In this last section, there are four long pauses. The speaker has stated at the beginning of the section that she is 'not very sure' about her uncle's place of origin. The abundance of pauses and especially long pauses may reflect this uncertainty. The pauses could also be seen to 'divide off' the speaker's opinion from her uncle's opinion. Long pauses occur after the phrase I thought it was Perth, and before I thought they came from Doune, thus surrounding the units which refer to the uncle's opinion.

Long pauses in this extract of Glasgow English, therefore, seem to function in exactly the same way as long pauses in Edinburgh English, i.e. they delimit paratones. The shorter pauses which measure 0.3-0.54 secs function not only as boundary pauses but also as planning pauses in this extract (cf. examples 1 and 7 already

quoted). Thus because the boundary pauses are shorter than the boundary pauses observed in the Edinburgh extract, there is no difference in length between boundary pauses and planning pauses. How then are we to identify the boundaries of tone groups? We must have recourse to internal criteria.

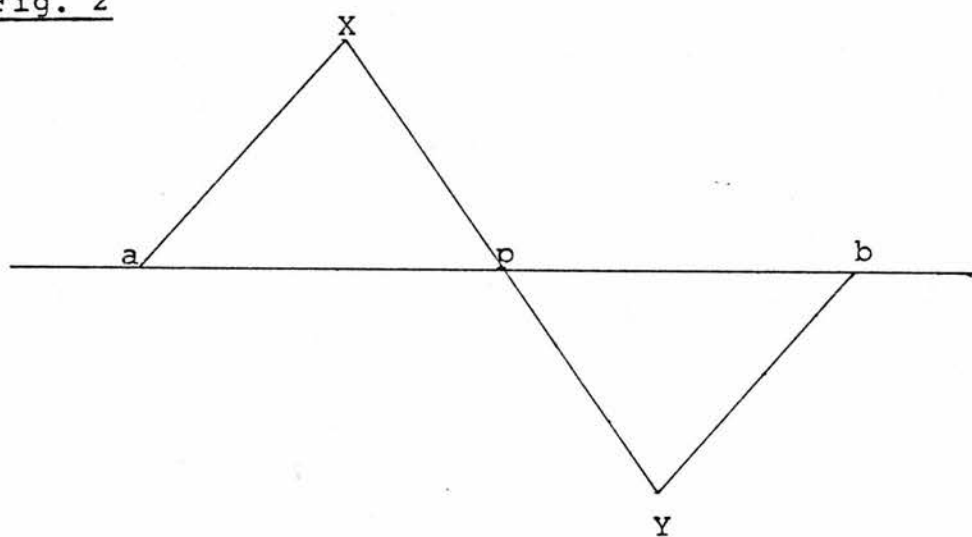
Let us examine the shortest pause-defined units which are greater than one word long (the stressed syllables are underlined):

1. he's due in +
180 160 160-180
4. for three or four weeks +
160 225 170 160 150-170
9. the way he spoke
180 250 180 170-185
12. I thought they came from Doune +
200 180 170 180-160 160-200

These units are strikingly different from the units observed in the samples from Edinburgh and Thurso. The feature common to all of the above units is the fact that each unit ends with a rise on the final stressed word of the unit. Any preceding stressed syllables are realised on peaks of prominence in a similar way to the peaks of prominence observed in the Edinburgh sample, but only when there are unstressed syllables between the final stressed syllable and the preceding stressed syllable (as in units 9 and 12).

In units 1 and 4 there is a stressed syllable immediately preceding the final stressed syllable, and this penultimate stressed syllable is realised on the same level as the beginning of the final rise. Stressed syllables in Glasgow English do not deviate consistently from a baseline in the way that stressed syllables behave in Edinburgh English. If we take unit 9 as an example of a contour unit with two stressed syllables surrounded by boundary pauses, the contour unit in Glasgow English is realised as follows:

Fig. 2



The contour unit in Glasgow English is similar to the contour unit in Edinburgh English in that there is regularly more than one 'prominent' item in each unit. They are also similar in that there seems to be consistent 'aiming points' in both accents, the main difference between the two contour units being that in Glasgow

English, the final 'peak of prominence' is 'reversed', with the result that the final stressed element is low and not high. Because of this basic difference, Glasgow does not seem to possess a baseline in the same sense that Edinburgh does. In Edinburgh, the pivot point was the most consistent point of the contour unit, but the pivot point in Glasgow English can be realised anywhere between the two main aiming points due to a 'sandhi' effect which draws the unstressed syllables towards the stressed syllables. Peaks of prominence in Edinburgh English are perceived as prominent because they deviate upwards from an established baseline; peaks of prominence in Glasgow English are perceived as prominent because they deviate either upwards or downwards from a 'reference line'. The final stressed syllable of a contour unit will normally be realised as 'low' and will rise back to a 'reference line'. Preceding stressed syllables may be realised as 'high' (as in units 9 and 12) or as 'low' (as in units 1 and 4).

Let us examine longer stretches of speech which are delimited by pauses. The stressed syllables are underlined:

2. oh no they lived in Canada eh he
 190 150 160-180 180 180 180-160 160 200
- was married to my mother's sister +
 160 150-180 160 170 200-190 150-160

4. but he's coming home they're coming home
 160 180 170 160-175 180 150 180

on Sunday +
 160 150-180

7. he was born in the Highlands and he wants
 160 160 180 175 175 165-175 200 200 160-200

to go back +
 180 180 160-180

8. now I'm not very sure I thought it was
 200 220 250 200 170 190-220 180-160 160 160

Perth +
 160-185

If the above units are subdivided into tone group length with only two or three stressed syllables in each tone group, then the contour shape as exemplified in Fig. 2 is realised consistently for the following units given in brackets:

2. (oh no) ... (eh he was married) (to my mother's sister)+
4. (but he's coming home) (they're coming home on Sunday)+
7. (he was born in the Highlands) (and he wants to go back) +
8. (now I'm not very sure) (I thought it was Perth) +

The only group of stressed syllables which does not conform to the contour outlined in Fig. 2 is the unit (they lived in Canada). This unit contains an unstressed syllable which rises up to 180 cps before the first stressed

syllable. The first stressed syllable, the following unstressed syllable and the second stressed syllable are all uttered at the level of 180 cps. There is then a fall during the final unstressed syllables. This unit demonstrates the irregularity of the unstressed syllables in Glasgow English. Instead of returning to a reference line, the unstressed syllable is drawn towards the level of stressed syllables surrounding it. In this case, the unstressed syllable occurs at exactly the same level as the stressed syllables on either side.

Units 2, 4, 7 and 8 illustrate the fact that tone groups are combined into sense groups in exactly the same way in Glasgow English as described in Edinburgh English, i.e. there is a general tendency for pairs of stressed syllables to exhibit a typical contour pattern in each accent. This pattern is then repeated linearly to form 'chunks' of speech larger than the tone group. This larger unit is surrounded by boundary pauses, and will be referred to as a sense group (see also sections 7.1 and 7.2).

This sample of speech from Glasgow therefore exhibits the same patterns which were observed in the Edinburgh sample. Stressed syllables combine to form tone groups, the unmarked tone group in each accent forming a typical contour shape. In Edinburgh the stressed syllables deviate from a low baseline to form high peaks of promin-

ence. In Glasgow, the stressed syllables deviate from a mid reference line to form 'peaks' of prominence which are realised as high, above the reference line, or low, below the reference line. The normal contour shape for a tone group (which contains two stressed syllables) in Glasgow English will be in the form of Fig. 2, i.e. containing one high initial peak of prominence, one low final 'peak' of prominence, and a final rise which returns the speaker to his reference line.

Boundary pauses in Glasgow English are much shorter than the boundary pauses observed in Edinburgh English, or in the sample of Thurso English. It is therefore difficult to differentiate between pauses which are functioning as planning pauses, and pauses which are functioning as boundary pauses. The internal structure of the tone group in Glasgow English must serve to mark the domain of the tone group in many cases. Since the contour shape in Glasgow English typically ends with a final rise from low to reference line, the internal 'finality' marker operates fairly efficiently as a marker of tone group boundary. Long pauses function in all three samples as markers of topic shift.

7.4 Conclusion

Having examined samples of spontaneous speech from three different accent areas, it is obviously the case that

the systems operating in the text readings can also be applied to spontaneous speech. The pause system as outlined in chapter 6 operates in spontaneous speech, and the contour types which were realised systematically in the texts were also found in the sample of spontaneous speech from Edinburgh and Thurso. Glasgow English also exhibited a typical contour pattern, but this pattern differed from the pattern proposed for Edinburgh English and Thurso English.

Throughout the samples of spontaneous speech performance variables were seen to be very much in evidence. Planning pauses, hesitation, repetition, incomplete units, all served to interfere with the rather uncomplicated system proposed for the text readings. It was also the case that emotional involvement and interaction strategies interfered with this uncomplicated system proposed in chapter 6. Thus although the same systems can be seen to underlie spontaneous speech, the underlying system is often obscured in spontaneous speech due to performance variables and interaction strategies.

Appendix 1Spontaneous speech: 22BS; Edinburgh

(Pauses of less than 0.3 secs are denoted by +; pauses between 0.3-0.85 secs are also denoted by +, and the length of pause is given in brackets; long pauses are denoted by ++ with the measurement given in brackets; Fo measurements are given under the appropriate words of the text with dashes joining measurements which occur on the same word; interviewer's comments are given in brackets)

(what do you regret about the South Side?)

1. I regret + putting the people out of the
155 200 150 115 190 160 120 115

out of the South Side and central
120 105 105 120-110 120-105 110 120-115

Edinburgh you know + (0.86 secs)
120-100 105 140-100
2. I don't think ++ (1.8 secs)
145 175 100
3. especially after the war you know after
165-115 140-110 110 150 110 120 150

the + (0.64 secs)
100
4. war when they started the + (0.32 secs)
120 120 100 130-100 100
5. redevelopment and the ++ (1.00 secs)
105 120 100-120 110 110
6. well the authority more or less made it that
120 135 165-100 120 100 100 140 100 100

everybody was to go outside you know +
 125-100 100 100 100 125-95 100 130-100
 (0.68 secs)

7. the gardens and houses but + (0.3 secs)
 90 130-110 100 120-90 120

8. I would reckon that eighty per cent of
 160 140 120-100 100 140 140-100 120
 the people + (0.78 secs)
 100 120-90

9. didnae want to go out of the town
 150-120 160 100 100 100 100 100 125
 didnae want their gardens they were quite
 100 120-95 110 115-110 110 110 130
 happy where they were if they'd built houses
 130 100 110 120-100 125 115 125 100
 in the town (mhm) + (0.82 secs)
 150 - 105

10. which they're doing now after + (0.34 secs)
 160-125 160-145 160-120 165 115 110

11. about eh + (0.86 secs)
 110 105

12. (mhm) you know + (0.38 secs)
 160 120

13. thirty years too late so to speak
 145-120 125-115 140 120-150 130 80 110-95
 you know (mhm) + (0.62 secs)
 100 135-110

14. that's what I regret especially central
 165-180 110 110 120-90 155-110 120-115
 south Edinburgh + (0.8 secs)
 115-110 115-130-85

15. em ++ (1.5 secs)
 140-115

16. I would I reckon with taking the people
 110 125 160 180 120 200 120 120

out they've + (0.36 secs)
140-120 125-120

17. they've lost the community + (0.82 secs)
125 165 110 130-140-90

18. no community spirit at all I don't
180 130-90 110 110 120 120 180

think you've the same spirit in the new
140 100 100 130 125 100 100 125

housing schemes + (0.7 secs)
110-90 110-90

19. as you had in the old eh + (0.6 secs)
100 110 115 90 90 120 100

20. (did eh...) type of dwellings you know (mhm) ++
110 100 110-100 100 130-100

(1.8 secs)

Appendix 2.Spontaneous speech: 57IT; Glasgow

(Pauses of less than 0.3 secs are denoted by +; measured pauses are given in brackets as they occur in the stream of speech; Fo measurements are given under the appropriate words of the text with dashes joining measurements which occur on the same word; the interviewer's comments are given in brackets. Since it is difficult to differentiate between planning pauses and boundary pauses in this particular extract, units are numbered only when they form a 'cognitive whole'. The units in this extract will therefore resemble sense groups rather than tone groups).

1. em + my uncle's coming home from
 190-180 190 150-200 180 190-180 180
 Canada + (0.5 secs) + on + (0.46 secs) +
 180-200 160-150
 Sunday + (0.54 secs) + he's due in + (0.52secs)
 180-210 180 160 160-180
 (how long has he been away for, or has he just been
 away?)
2. oh no they lived in Canada eh he
 190 150 160-180 180 180 180-160 160 200
 was married to my mother's sister + (0.3secs)
 160 150-180 160 170 200-190 150-160
3. well she's been dead for a number of
 170-150 180 150 190-160 160 170-160 160

years now and he's remarried +(0.86 secs)
170 160-180 160 180 160-180

4. but he's coming home they're coming home on
160 180 170 160-175 180 150 180 160

Sunday + (0.28 secs) for three or four weeks+
150-180 160 225 170 160 150-170
(0.82 secs)

5. so we hope to see them while
200 200 210 160 150-160 155 160

they're here + (0.42 secs)
160 160-180

6. he wants to go back up to the
190 210 180 170 165 160 160 160

Highlands + (0.32 secs)
160-190-180

7. he was born in the Highlands and he
160 160 180 175 175 165-175 200 200

wants to go back + (0.32 secs)+ and see +
160-200 180 180 160-180 150 175-160

(0.42 secs)+ the Highlands while he's at
175 145-175 180 170 170

home + (0.42 secs)
160-200

(whereabouts in the Highlands do you know?)

8. now I'm not very sure I thought it
200 220 250 200 170 190-220 180-160

was Perth +(0.78 secs)
160 160-185

9. but + (0.48 secs)+ the way he spoke + (1.5 secs)
175 180 250 180 170-185

10. he seemed to feel it was further North+
180 200 180 175 200 200 180-175 170-225

(0.46 secs)

11. but he maybe considers + (0.3 secs)+
180-200 300 200-160 175-170-210

Perthshire + Doune you know (0.68 secs)
200-175 180-225 240 250

12. I thought they came from Doune+ (0.68 secs)
200 180 170 180-160 160-200

(that's Perthshire)

Chapter 8.

8.0 Introduction

There have been two main schools of thought in the field of intonation:

1) the British school, who tend to favour what might be referred to as the 'tonetic stress marks' system. This is the kind of system used by Kingdon (1958), O' Connor and Arnold (1973), Crystal (1969), and Halliday (1970). As the main system adopted by British investigators working in the field of British intonation, it was involved in the discussion of the theories of two of the above authors in chapter 3. That discussion was concerned mainly with the realisation of these systems. In this chapter, however, I propose to summarise briefly the more fundamental and general problems involved in the tonetic stress marks system.

2) the other main school of investigators working in intonation is composed mostly of Americans working with the 'pitch levels' system. This system is the one used by Pike (1945), Trager and Smith (1951), Harris (1944) and Hockett (1955 and 1958), inter alia. Some of the general problems involved in the latter system will also be presented briefly in this chapter. I will then outline the theoretical approach advocated in this thesis, and present support for such an approach.

8.1 The tonetic stress marks system

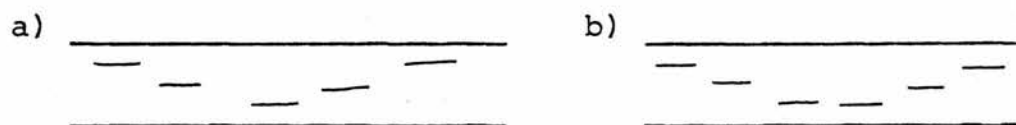
The basic tenet of this theory is that the intonation pattern of an utterance is built on a framework consisting of stressed syllables. Each stressed syllable can have a particular tone on it (such as rising, falling or level) and when the tonality of the stressed syllables has been specified, the intonation pattern over the unstressed syllables is said to be predictable. We can go even further and say that once the tonality of the single 'nuclear' stressed syllable, the 'tonic', has been specified then the intonation over the whole tone group or unit can be predicted. Such a model would seem to be a very powerful one, but this position of apparent strength is gradually eroded by allowing for contrastive oppositions on pre-nuclear stressed syllables. Some theories also allow for double tonics, thus permitting two positions of distinctive oppositions within a tone group. Most of the analysts mentioned above go even further and recognise at least a two way distinctive opposition operating within the unstressed syllables in the pre-head position (e.g. between 'high' and 'low'; cf. O'Connor and Arnold (1973), p. 22-7; Crystal (1969), p. 147; Halliday (1963), p. 11). The tonetic stress marks theory rapidly approaches the point where there can be said to be no intrinsic restriction of the number of tones, nor any restriction upon the distinctions that can be

expressed either by the tones themselves, or by symbols representing raised, lowered, etc.

The main weakness of this type of theory is the very fact that the theory deals with individual syllables and is unable to account for the relationships holding between syllables within a tone group in any structured way. The tonetic stress marks theory is also unable to constrain the number of patterns which can be specified by the system. New tones and new symbols seem to be introduced in ad hoc ways to account for any pattern encountered which is not captured by the existing framework.

The fact that each syllable is taken into consideration leads Crompton (1978) to point out that such a system cannot account for what he would like to call 'spreading' (p. 6), nor can it account for the similarity between such patterns as the following in Fig.1:

Fig. 1



Both of the above patterns have exactly the same Fo contour, the only difference being that the trough is spread over two syllables in lb). Crompton would suggest that the tonetic stress marks system fails to capture this

'identical' intonation pattern and is forced to treat both patterns as distinctive. This means that the tonetic stress marks system has no way of constraining the number of intonation patterns which can be expressed by the system, and also fails to express only those oppositions which operate distinctively in the language.

8.2 Pitch levels system

This theory is based on the belief that intonation patterns can be represented in terms of 'pitch phonemes'. Most of the investigators working in this field operate with four levels of pitch. These levels or 'pitch phonemes' are represented by numbers, with the highest number normally corresponding to the highest pitch level. These pitch phonemes are said to represent varying fundamental frequency values, since 'relative' pitch values are said to be distinctive in the intonation of an utterance, and not 'absolute' values. Again, this system operates on an utterance syllable by syllable. This theory does restrict the number of terms in the system very strictly (normally to four terms, 1, 2, 3, and 4), but it fails to capture the similarity common to all of the following contours; /21/, /31/, /41/, /42/, /32/, /43/. All of these configurations are variations of fall, but this generalisation cannot be expressed by the pitch levels theory. Thus, although the pitch levels theory seems to

be stronger than the tonetic stress marks theory because the number of distinctive terms is restricted, the pitch levels theory fails to express only those oppositions which operate distinctively in the language.

Because the number of distinctive oppositions is tightly constrained the reality of these oppositions can be tested, and the results show that a restriction to 4 levels seems arbitrary.

Lieberman (1965) examined the ranges of fundamental frequency values associated with the four pitch levels in the transcriptions of a set of test utterances carried out by two linguists. He found a considerable amount of overlap between the ranges. He says:

"The Trager-Smith pitch levels do not correspond to discrete non-overlapping ranges of fundamental frequency nor do they correspond to discrete relative ranges of fundamental frequency. These comments apply even when we consider the transcriptions made by a single linguist who carefully transcribed the tape recorded sentences of a single talker."

(Lieberman, 1965, p. 47)

At best, then, the pitch levels system does not specify the relationship between the fundamental frequency and the four pitch levels.

The investigators working on British intonation using the tonetic stress marks system constantly use more than four distinctive levels of pitch in their transcriptions. Kingdon (1958, p. 25-9) and Crystal (1969, p. 225-233) consistently use six and seven distinctive levels. Does this mean that English has 6 or 7 pitch levels instead of 4? Presumably this is not the case, since other examples could probably be discovered requiring even more distinctive levels. Perhaps the levels should not be seen as a restricted number, but should be treated as notational devices for expressing the height of one syllable relative to another. Bierwisch, for example, has done precisely this. He abandons the fixed number of levels notion, preferring a system which uses numbers to express the height of one syllable relative to another (cf. Bierwisch, 1966).

Investigators using the pitch levels system claim that intonation is essentially a matter of reaching particular levels at particular points in time, thus implying that pitch movements such as rise and fall versus jump or step are merely means to this end. Note Pike (1945) who states .. "In any rising or falling contour, two contour points are present: the pitch level at its beginning and the pitch level at its end." (p. 27) The pitch levels system therefore differs fundamentally from the tonetic stress marks system which claims that pitch movement direc-

tion is the essential feature. Each system, then, is structured to capture the generalisations relating to the fundamental beliefs upheld by the system itself and each system fails to capture generalisations not catered for by the fundamental claims of the theory.

In summary, the tonetic stress marks system deals essentially with distinctive oppositions between directions of pitch movement and not with distinctive oppositions between levels of pitch height on the whole. The system is expressed by a transcription apparatus which notes the height of one syllable in relation to another in terms of a variable spatial relationship between two lines, thus not restricting the number of levels at all. The pitch levels system restricts the number of levels to a small fixed number, but does not express generalisations about pitch movement (as noted earlier in this section). Both of the above theories operate in terms of single syllables and relate each syllable to its neighbour in such a way that generalisations about contour configurations cannot be expressed by either theory.

I would like to propose a general theory which would account for the relationship between one pitch height and another; which would express generalisations concerning the direction of the pitch movement; and which would capture generalisations about the contour configurations of various intonation patterns.

8.3 The proposed model for Scottish English

As I stated in chapter 4 (section 4.2) the intonation pattern of the data was transcribed using a stave system similar to the ones used in tonetic stress marks systems.

A three-line stave^{was used} for transcribing which could then be converted into 5 levels (cf. Fig. 1, p. 109). These 5 levels formed the basis of a number system which could be used as an alternative transcription system. We noted in chapter 4 that these different methods of transcription, stave versus number system, produced different outputs.

The stave was used as a visual representation of the various pitch heights reached during the course of an utterance. The limits of the stave (especially the upper limit) were rarely used. The stave also placed no restriction upon the number of contrastive levels of pitch height which could be represented, whereas the impression of movement could be expressed fairly easily. The stave system of transcription can be seen to resemble closely the tonetics stress mark system analysis.

We also used a number system for transcription. We used five levels because we had already discovered through using the stave system that a 'mid' line of some sort was an essential reference point for transcription purposes, i.e. it seemed to be perceptually easier to set up a mid-type reference line from which syllables departed in both directions, than to set up perceptual boundary lines

defining the upper and lower limits of a speaker's range. The five level number system was therefore set up to correspond to the stave system as closely as possible. The number of contrastive levels of pitch height could only be captured in a very general way since /35/ could represent either a rise which is realised as continuously moving pitch, or as a jump from level 3 to level 5. We tried to differentiate between these two different realisations by linking the numbers to represent continuous movement (i.e. /3 - 5/ and by separating the numbers to represent a jump in pitch (i.e. /3 5/). These two alternatives would have to be treated as the extremes of a varying continuum, however, since intermediate variations of pitch movement are possible.

Each system captures different features of the pitch characteristics of an utterance, therefore both systems were used simultaneously where appropriate. Depending on the context of the analysis, each system could also be used individually. For example, if we wished to see how the pitch movement of a certain type of contour was realised in different phonetic contexts, we might use only the stave system. If we wished to examine the relationships between different peak heights within a contour, we might use only the number system. If we wanted a detailed representation of a particular contour we would use both systems simultaneously. Both systems were therefore

available for use either together or separately depending on the context of the analysis required.

The stave system and the number system are both descriptive systems of course, and as such have nothing to say about the distinctive oppositions in operation. The systems of analysis involved in intonation operate at various levels, each level differing from the other in degree of abstractness. We will now discuss the various analytical systems which operate in intonation.

8.3.1 The fundamental system of intonation

The most 'basic' level of intonation analysis which can be said to contain distinctive oppositions and thus constitute a system, is the level containing the STRESS vs. UNSTRESS opposition. This level can be said to be the most fundamental level of intonation. There is only a two-way distinction in operation (+ stress) but this distinction forms the basis for the other systems of intonation which I am about to propose. Any utterance which has had its stressed elements identified can then be allotted its basic intonation pattern (I will use 'pattern' to describe the eventual configuration of an intonation contour). While these elements, stressed and unstressed, are being proposed as the basic underlying core features of all intonation patterns, they are in fact frequently realised as a two-way distinction at the

surface level in actual utterances in the data. In the sample of Scottish speech from Edinburgh, Thurso, and Lockerbie studied for this thesis, this system of intonation is realised in such a way that the stressed elements are 'high' and the unstressed elements are 'low'. The phonetic environment will determine whether the transition between low and high is realised as a jump in pitch or a glide in pitch. Note Isacenko and Schadlich (1970) who state that:

"It is possible to represent the typical intonation patterns of the German sentence as a sequence of two and only two pitch levels. In our experiments discrete utterances were put together with different fundamental frequencies in such a way that there was no glide in the signal. These abrupt changes from one pitch to another are perceived by the ear as transitions and cannot be distinguished from natural glides from one pitch to another." (1970, p. 57)

Since the stress system is at present a syllable based system, this means that the fundamental system of intonation is also 'syllable' based. The following utterance, for example, might have the underlined elements as the stressed elements: 'he thinks they should really do something about it'. These stressed elements would then be assigned the appropriate values in the fundament-

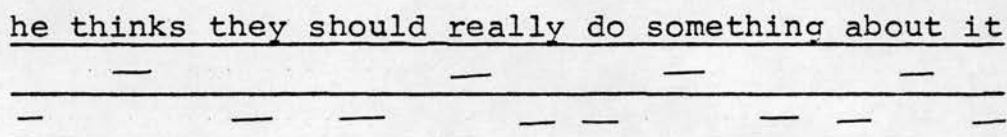
al intonation system as follows:

he thinks they should really do something
 low high low low high low high low
 about it
 low high low

This level of intonation analysis, then, can be seen to be very close to the basic rhythm of the language. I would like to argue that the two systems (the rhythm system and the fundamental system of intonation) correlate so closely that, at the fundamental level of intonation analysis (i.e. following stress assignment), the two are very difficult to define separately. The fundamental system in intonation depends solely on a two-way opposition between stress and unstress, and is realised as a two-way opposition between 'low' and 'high' pitch height (in Edinburgh, Thurso and Lockerbie at least). Stress (as pointed out in chapter 1) relies heavily on pitch as a cue for its identification acoustically, but involves many other parameters such as intensity, duration, and vowel quality which are not involved in the fundamental intonation system. The fundamental system of intonation is an abstract system. If this system is realised in its simplest form at the surface level, it is identified according to the pitch parameter alone.

The two values high(H) and low(L) can be regarded as

the basic departures from our transcriptional reference line. The above example would be realised on the stave as follows:



We might therefore consider (H) and (L) to be the two fundamental pitch phonemes (to borrow a term from the upholders of the pitch level system). It should be noted that these two terms, as used in the present system, are abstract. No claims are being made about the relationship between these terms and the fundamental frequency ranges which might be involved in their realisation. In the number system which we adopted for transcription purposes, the above example would be realised as follows (again no implications about the frequency range which might realise these abstract levels):

he thinks they should really do something about it
 2 4 2 2 4 2 2 4 2 2 4 2

At this fundamental level of intonation, the system does not require to delimit utterances in any way, (L) and (H) can be assigned to an utterance of any length.

8.3.2 The tone unit system

The next level of analysis above the identification of stressed and unstressed elements is concerned with how these elements combine to form basic tone units. The tone unit is also an abstract unit in this analysis. The tone unit contains only one stressed syllable, all preceding unstressed syllables and any following unstressed syllables preceding silence, or belonging to a polysyllabic ^{word} which contains the stress. Consequently, each tone unit will contain only one stressed syllable and a variable number of unstressed syllables. This unit can be compared with the 'foot' in the rhythm system (cf. Abercrombie, 1967).

The above example would be allotted tone units as follows:

he	<u>thinks</u>	they	should	<u>really</u>	do	<u>something</u>	<u>about</u>	it
2	4	2	2	4 2	2	4 2	2 4	2

8.3.3 The tone group system

The third level of analysis is concerned with how tone units are combined to form tone groups. The tone group may contain any number of tone units, and the relationship between the stressed syllables of the tone group will be determined by the context of the utterance and the speaker's intent. The relationship between these syllables cannot be predicted at the level of tone group

assignment. The tone group, then, could have as many 'tonics' (in Halliday's system) as it has tone units, or stressed syllables (see chapter 5). Each stressed syllable will normally be realised as a peak of prominence, but this peak may not always be perceived. Some stressed syllables may be boosted and some may be depressed. Thus peaks occurring between boosted peaks may be ignored^{by the Hearer} and some depressed peaks may not be large enough to be perceptually relevant. Whenever the stressed syllable is perceived as a peak of prominence we will refer to these stressed syllables as 'phonological foci'. When a tone group has only one phonological focus, or when a phonological focus is 'boosted' for some reason (e.g. in the context of lexical contrast) then this single, easily identifiable, phonological focus could be referred to as the 'tonic' of an utterance. Thus tonic in the present system would conform to the requirements of both Halliday and Crystal who state that there is only one tonic in a tone group. The qualification on tonicity demanded by the present system is that if and when there is only one phonological focus in a tone group, then it may be referred to as a tonic. The present system does not require that there be only one tonic in every tone group. On the contrary, it would seem to be the exception rather than the rule in the data examined for this thesis (cf. chapter 5).

The unmarked (or most common) tone group in the typ-

es of English mentioned in 8.3.1 consists of two tone units, with two phonological foci which are realised as two equal peaks of prominence, one near the beginning of the tone group and one near the end. In slow unhurried speech (as for example in the readings of the text) the tone group will be delimited by a boundary pause, normally lasting in the region of 0.6-0.8 secs.

8.3.4 The sense group system

The sense group is a unit^{realised} in the intonation system which is defined essentially in syntactico-semantic terms. The sense group is a unit of intonation which can be correlated with the clause in the syntactic system. The sense group will frequently contain more than one tone group, although it may be co-terminous with the tone group. The domain of the sense group will normally be greater than the tone group and it will normally be delimited by a boundary pause (cf. chapter 7).

An examination of the data showed that the tone group did not always co-occur with the syntactic unit 'clause' and, in fact, the two units sometimes mis-matched in such a way that confusion and ambiguity resulted. It was therefore decided to separate the two levels of analysis, such that the tone group is defined in phonological terms, and the sense group in syntactico-semantic terms. The above

(p415)
 example therefore could be analysed as 2 tone groups
 and 1 sense group.

8.3.5 The paratone system

There are two different types of paratone as mentioned in chapters 4, 6, and 7 (specifically section 7.1.2), but both are structured in the same way. A paratone will consist of several sense groups. The sense groups within a paratone will be related thematically in one of two ways, thus differentiating between the above two types of paratone. The tone groups may be linked grammatically (giving a paratone with cohesion) or purely thematically (giving a paratone with coherence). The first tone group of a paratone will be boosted such that the whole tone group is raised (or expanded) in pitch range. The first stressed lexical item of the first tone group will normally be especially high. The pitch range will gradually return to normal during the next one or two tone groups. At the end of the paratone, the tone groups will become reduced in total pitch range until the last tone group of a paratone which will normally be realised on the lowest pitch range of the paratone. There may then be an optional 'summarising' sense group, which may be almost as high as the first tone group of the paratone. This 'summarising' sense group will contain a summary of the theme of the paratone and function as a 'rounding-off' of one topic, thus preparing the way for a new topic.

The paratone will normally be bounded by pauses which are longer than the pauses surrounding the tone groups (i.e. in the region of 1-2 secs).

8.4 The realisation of the model

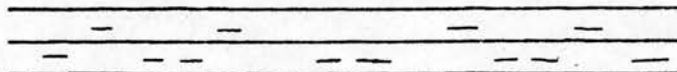
The above section describes 5 different systems each operating at a different level of abstraction. These 5 systems can be seen to operate on a basic output from rules assigning stress. The configuration of the intonation pattern develops from the fundamental level where the intonation is closely interrelated with the stress system, to a close-approximation of the realisation of an intonation contour which can be described in either of the transcription systems used in this thesis.

The following figures demonstrate how the various systems contribute to the development of the pattern:

1) FUNDAMENTAL SYSTEM	L H L L H L L H LL H L
2) TONE UNIT ASSIGNMENT	(L H) (L L H) (L L H) (LL H L)
3) TONE GROUP ASSIGNMENT	((L H) (L L H)) ((L L H) (LL H L))

which might be realised as-

or as-

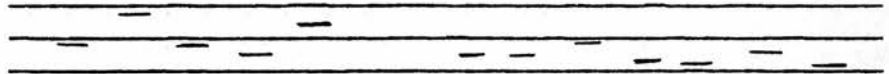


4) SENSE GROUP +((2 4 2 2 4) (2 2 4 2 2 4 2))+
ASSIGNMENT

5) PARATONE
ASSIGNMENT

which might be
realised as- ++(3 5 3 2 4) ... (2 2 3 1 1 2 1)++

or as-

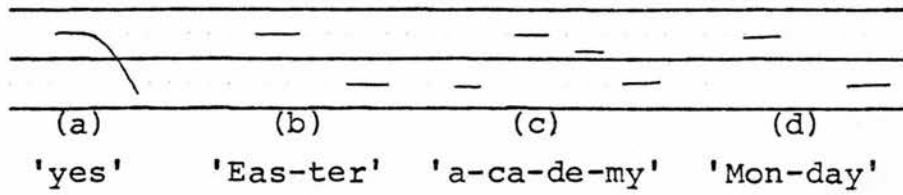


Thus by specifying a basic contrast (H) vs. (L) which is realised on stressed and unstressed syllables respectively, and by applying fairly simple rules to this basic contrast we can approach a form which is very close to the surface forms collected in the data. This theory can be seen to account for the relative height of groups of syllables quite successfully. It might be argued that such a theory does not successfully capture generalisations about the direction of pitch movement. I do not think this is true.

The closer we draw to the realisation forms, the more we are concerned about the configuration of the contour as a whole and how the contour as a whole operates contrastively. For example, a tone group with a question function will either rise overall, or be raised to high in the overall pitch range, or have an extra-high stressed syllable. All of these configurations seem to mark the function 'question' in Scots (cf. also Studdert-Kennedy and Hadding-Koch's (1973) results.

If we are taking into account the total configuration of a contour we must be concerned with the overall undulation of the fundamental frequency and its auditory correlate, pitch. The realisation of the pitch movement will depend on many different features of the phonetic environment. The presence or absence of voiceless segments will affect the realisation of pitch movement. If there is a cluster of voiceless segments occurring between two vowels, there will be a jump in pitch instead of continuous pitch movement (see Fig. 2b). The number of unstressed syllables occurring between peaks of prominence and other phonetic factors will affect the realisation of the pitch movement. In section 4.2, it was pointed out that when the final stress occurs on a monosyllable and there are no unstressed syllables following it, then the stressed syllable will be realised on a continuously moving fall or rise, unless of course the syllable is level. If there are unstressed syllables following the stressed syllable, the pitch movement will be distributed across the syllables (see Fig. 2c). If the stressed syllable is a short vowel followed by one unstressed syllable the pitch differential will be realised as a jump in a similar way to Fig. 2b (see Fig. 2d):

Fig. 2



All of the above configurations in Fig. 2 are realisations of the same contour. This contour could be described as (2)-4-2, meaning that the contour is a fall of a neutral type with an optional space for an unstressed syllable preceding the fall. The realisations of the contour vary because of the phonetic characteristics of each utterance.

Generalisations concerning the direction of pitch movement must be expressed if we are examining the contour as a whole. The overall general configuration which can be abstracted from the various realisations of a contour will form the basis for 'contour-types' such as those mentioned in chapter 6. These contour types will then operate contrastively according to their position within the paratonic structure (as exemplified in chapters 6 and 7), and according to their discoursal, illocutionary and attitudinal function.

Let us examine a sample of data, using the number transcription system, in order to see what kinds of problems arise when we attempt to analyse real data in terms of the above levels of abstract analysis.

The data which we will examine was transcribed as

follows:

+ until the war started in nineteen thirty nine you
 2 4 3 5 4 2 2 4 2 3 2 4-2 2
 got an extra haep'ny which made it one and
 2 2 4 3 5 2 2 4 2 3 2
 eightpence haep'ny +
 3 2 5 2

One feature which can be noted immediately is the effect of tonal sandhi, i.e. when an unstressed syllable occurs preceding a boosted peak (i.e. level 5) and following a normal peak (i.e. 4), the unstressed syllable is boosted to 3. Thus an abstract contour type (4 - 2 - 5) would be realised as (4-3-5) -- cf. until the war and extra haep'ny as opposed to eightpence haep'ny where the preceding peak is level 3 and not level 4.

If this data is analysed at the first level of analysis (the fundamental level) the analysis would be as follows:

+until the war started in nineteen thirty nine you
 L H L H H L L H L H L H L
 got an extra haep'ny which made it one and
 L L H L H L L H L H L
 eightpence haep'ny +
 H L H L

The second level of analysis would assign tone units as follows:

+ until / the war / started / in nineteen / thirty /
 nine / you got an extra / haep'ny / which made it /
 eightpence / haep'ny / +

The third level of analysis would then assign tone group divisions to the above sample of speech, but there are no pauses within the sample, only pauses surrounding the sample. Does this mean that the whole sample has to be created as a single tone group? If we take the context of this conversation into account, we would find that this conversation is not slow, unhurried and free from possible interruptions. On the contrary, the speaker is speaking very quickly about a subject which his companions are only too willing to interrupt in order to give their own views on the subject. It is reasonable to suppose that in this situation, the speaker chooses not to pause after every tone group. Therefore in order to divide up the stream of speech into tone groups we must have recourse to the internal criteria which typify or define a tone group, i.e. into units which contain two peaks of prominence, one near the beginning of the tone group, and one near the end. Tone groups might therefore be assigned as follows:

+ until the war started / in nineteen thirty nine /
 2 4 3 5 4 2 2 4 2 3 2 4-2
 you got an extra haep'ny / which made it (/)
 2 2 2 4 3 5 2 2 4 2
 one and eightpence haep'ny / +
 3 2 3 2 5 2

Problems arise at this level of analysis when we are dealing with 'real' speech. The first tone group contains two peaks of prominence and is straightforward enough, but the second tone group presents slight problems in that there are three peaks of prominence. Since the three peaks occur on one lexical item (the numeral 1939), it would seem more logical for the analyst to assign one tone group to this numeral than to split it into more than one tone group. If we examine the contour pattern, we find that it conforms to one of the contour-types proposed in chapter 6, i.e. an equal peak contour with intervening pivot peaks (see Fig.26). Thus the configuration is in fact a variation of the equal-peaked contour, contour XY.

The third tone group presents no problems, again we have two peaks of prominence which typify the internal features of the tone group. The remainder of the sample however could be analysed in three different ways a) we could extend the argument used for the second tone group since again, we have a tone group containing

two peaks of prominence, one near the beginning on the item made and one near the end of the tone group on the item haep'ny. There are intervening pivot peaks on the elements one and eight. Thus we seem to have a contour of the type XY. b) alternatively one could posit the existence of two tone groups each containing two peaks (i.e. /which made it one and/, and /eightpence haep'ny/). However one might argue against the latter proposal on the grounds that the sum of money (one and eightpence haep'ny) should not be divided between two tone groups but should be kept as a single unit. (c) The argument proposed in (b) might be extended to posit the existence of the two tone groups /which made it/ and /one and eightpence haep'ny/, which splits up rather a large 'tone group' yet doesn't divide up the sum of money which should perhaps be treated as a single unit at this level.

All of the above analyses are possible, but we must decide which analysis is the most probable one. If we have recourse to notions of sense, then I think analysis (b) must be discounted on the grounds that a sum of money which is functioning as a single lexical item should not be split between two tone groups. If we then examine the analyses (a) and (c), we might suggest that these different analyses are in fact analyses at different levels, such that we might prefer to adopt analysis (c) in order to assign tone groups, and analysis (a) to assign sense groups;

the sense group assignment, then, might be as follows:

+ until the war started in nineteen thirty nine /
 you got an extra haep'ny / which made it one and
 eightpence haep'ny / +

The first four levels of analysis could therefore
 be applied to the above sample of speech as follows:

UNTIL	THE	WAR	STARTED	IN	NINETEEN	THIRTY	NINE	YOU
L H	L	H	H L	L	H L	H L	H L	L
_____TU_____		_____TU_____		_____TU_____		_____TU_____		_____TU_____
_____TG_____				_____TG_____				_____TG_____
_____SG_____				_____SG_____				_____SG_____
GOT	AN	EXTRA	HAEP'NY	WHICH	MADE	IT	ONE	AND
L	L	H L	H L	L	H	L	H	L
_____TU_____		_____TU_____		_____TU_____		_____TU_____		_____TU_____
_____TG_____				_____TG_____				_____TG_____
_____SG_____				_____SG_____				_____SG_____
EIGHTPENCE	HAEP'NY							
H	L	H L						
_____TU_____		_____TU_____						
_____TG_____		_____TG_____						
_____SG_____		_____SG_____						

(each underlined element represents the extent of
 the unit defined in capitals as follows: TU= tone unit;
 TG= tone group; SG= sense group. The unit divisions are
 marked in capitals below the relevant words)

The fifth level of analysis (i.e. the placing of
 this sample in paratonic structure) has to be carried out

with specific reference to the discourse structure in which it appears. The levels of analysis proposed so far can be defined phonologically (with the exception of the sense group), but as the various levels of analysis contribute to the development of the intonation pattern, there is a gradual progression out of the realm of phonology into the realm of discourse structure. In order to 'place' a particular sense group or series of sense groups within the structure of a paratone, we must know its place in the structure of the message.

The conversation surrounding the sample under consideration is given in Appendix 1, the conversation prior to the above sample being centred round a discussion about a bricklayer's pay. We are told that fast bricklayers were paid a halfpenny an hour extra to build up the corners of a building, thus encouraging the average bricklayers to work faster. The above stretch of speech is then uttered, but its paratonic function is unclear. Each of the three sense groups contains a boosted peak (realised as 5 in the number system). This means that this series of sense groups could either be functioning as a 'new' development of the theme (see chapter 7), or functioning as a 'summarising' series of sense groups (see section 8.3.5 and chapter 7).

If the first function is adopted, then the above sample would mean 'all of the previous conversation happened before the war, but after the war, the average brick-

layer got one and eightpence halfpenny an hour'. The hearer seems to prefer this meaning, but is not quite sure whether this is the intended meaning, since she asks: "and put the others up to one and nine, no?" i.e. if the average bricklayers were getting one and eightpence halfpenny, then the fast bricklayers must be getting one and nine. However, the hearer's comment is neither confirmed nor denied by the speaker.

The alternative meaning of this sample of speech if functioning as a 'summarising' unit might be that 'all of my previous conversation has contained the message that fast bricklayers were paid one and eightpence halfpenny an hour which was a halfpenny more than the average bricklayer; all of this took place before wartime'.

The hearer's comment which questioned her interpretation of the message was neither confirmed nor denied, with the result that the ambiguity was not resolved. The intended meaning of this sample is unclear, i.e.

- either a) fast bricklayers get one and ninepence
 average bricklayers get one and eightpence
 halfpenny (after the war)
- or b) fast bricklayers get one and eightpence
 halfpenny
 average bricklayers get one and eightpence
 (before the war)

The ambiguity is not resolved by the intonation because both functions (i.e. 'new' development of theme and 'summarising') are realised by the same phonetic markers, i.e. boosted peaks of prominence. When we deal with real speech, therefore, we find that systems can overlap and in fact result in ambiguity. It should normally be possible to differentiate between the two functions mentioned above by having recourse to criteria in other areas of the language. For example, a sense group which is functioning as a 'summarising unit' will normally consist of a deictic referent such as 'that' denoting 'all that has gone before', plus a summary of the theme. Thus examples of a normal summarising unit are:

So that's what I think of the South Side
 So that's what happened to me in the States
 That was more or less the trend of houses
 That's the way that they bribed bricklayers too you
 know.

Sense groups which are introducing new information, on the other hand, do not contain deictics such as that, do not contain a summary of the preceding theme, but do contain elements which have not been mentioned previously. Therefore normally the above two functions of an identical

phonological pattern would be differentiated on syntactic or semantic grounds. The phonological function of the above utterance is ambiguous because the syntactic structure and the semantic content do not disambiguate between the confused signals. Let us repeat the data below for ease of exposition:

+until the war started / in nineteen thirty nine /
 2 4 3 5 4 2 2 4 2 3 2 4-2
 you got an extra haep'ny / which made it / one
 2 2 2 4 3 5 2 2 4 2 3
 and eightpence haep'ny / +
 2 3 2 5 2

We do not have any deictic referent such as 'that' in this sample; we do have the introduction of a time element which has not been specified before, yet the content of this sample repeats the theme of the previous section of speech. The time preposition until and the pronoun you are potentially ambiguous for some speakers resulting in either of the following gloss meanings:

- a) 'all this happened until the war started in 1939.
 At that time things changed with the result that everybody got an extra halfpaenny, making the average wage, one and eightpence halfpenny'.
- b) 'therefore, before the war started in 1939, fast

bricklayers got paid an extra halfpenny, making the fast bricklayers' pay, one and eight-pence halfpenny'

Thus in the above sample there is semantic ambiguity, no deictic referent, the same semantic content as the previous theme, and an introduction of a new time element. This confusion of syntactic and semantic information results in paratonic ambiguity which cannot be resolved by the phonology, and, in this case, is not resolved by syntactic or semantic features.

The realisation of the proposed levels of intonation analysis when examined in samples of read data demonstrates that as the analysis progresses from one level to the other, more and more non-phonological factors must be taken into account. Levels 1, 2 and 3 can normally be defined having recourse only to phonological features, but levels 4 and 5 involve syntactic and semantic features as well as phonological features. The phonology alone will not always determine units at the later^t levels of analysis.

8.5 Support for contour system

Bolinger (1964) describes intonation in the following terms:

"Accentual systems involve more than singling out important words by accenting them... We tend to favour the two extremes of the sentence (or, in longer sentences, the two extremes of each relatively independent phrase or clause), as if to announce the beginning and the end. There may be intermediate accents, but they are less prominent. This gives the sentence the shape of a bumpy suspension bridge."

He then gives as an example the following sentence which approximates very closely to the contour types we have been describing in chapters 6 and 7.

snow generally comes early in October. The

This example shows a similar configuration to the neutral contour type proposed for Edinburgh Scots. Bolinger has produced a vast number of articles on intonation and is regarded as an eminent scholar in this particular field, his observations, therefore, should not be taken lightly, but rather be regarded as the culmination

of many years of study (see Bolinger, 1951, 1958, 1961, 1963, etc).

Lieberman is one of the data-based investigators in the same field, and he states frequently that intonation is perceived in terms of complete contours (cf. especially 1965). There seems to be ample support throughout the literature therefore, for analysing intonation in terms of contour units.

In Peck's examination of the acoustic parameters of American intonation (1969) he states:

"It is perhaps significant that in describing the acoustic data, long patterns, or tunes, proved more useful than atomic pitch levels or rise and fall tones." (p. 106)

In Peck's study,

"'tone groups' consist of a syllable with raised Fo plus all the syllables before it that have a rising Fo and all the syllables after it that have a falling Fo, the boundaries being pauses or valleys in the Fo trace after the perturbations caused by articulations have been removed."

Thus Peck's 'tone groups' would correspond very closely to the proposed tone units in the present study. Peck

arrives at an analysis of intonation distributed over three levels, the first level dealing with the variations of F_0 caused by the physics of the vocal tract; the second level dealing with tone groups or stress groups as described above; and the third level dealing with the 'larger patterns' in which tone groups (or stress groups) appear. The latter two levels correspond almost exactly to the present analysis of tone units and tone groups. We therefore have an analysis of American English which corresponds very closely indeed to the present analysis of Edinburgh English.

8.6 Contour systems in other languages

One of the most thorough investigations carried out in the field of intonation at the present time is the investigation of Dutch which is being carried out by Collier, 't Hart, Cohen and their associates at the Institute for Perception Research, Eindhoven. The main area of research explored by these investigators has been concerned with describing the melodic properties of utterances in terms of perceptually relevant pitch events. The aim of this approach has been to establish the relationship between the continuously varying fundamental frequency of the speech signal and the corresponding discrete perceptual units. 't Hart and Cohen (1973) state the underlying principle of their research as follows:

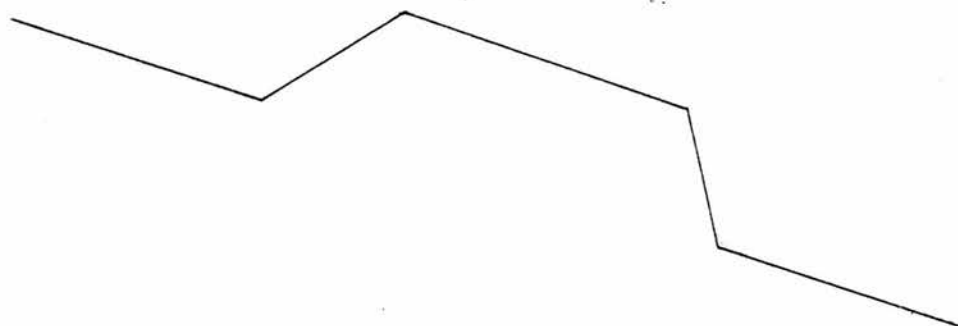
"... the effects of F_0 changes as such is no reliable measure for their perceptual relevance. These effects suggest that there is no one-to-one relationship between voice periodicity and the perception of speech pitch; rather, the listener interprets what he hears in terms of a limited set of recognisable patterns, which therefore may be assumed to possess a perceptual identity, and may even be considered perceptual units in some sense." ('t Hart and Cohen, 1973, p. 310.

The results of their research led Collier and 't Hart (1975) to propose that the stylised perceptual equivalent of a natural fundamental frequency curve should be called a pitch contour. The pitch contour "consists of a sequence of pitch movements which alternate between the higher and lower reference level of the so-called declination line." (p. 107) Their inventory of perceptually relevant pitch movements is as follows; declination line (equivalent to our proposed baseline); high and low (equivalent to the proposed pitch levels (H) and (L)); two types of rise, one with prominence the other without; two types of fall, one with prominence the other without; gradual fall; and half fall (these rises and falls being realised in the present system as peaks of prominence.

The above perceptually relevant pitch movements are

then seen to form a canonical pattern for the major class of Dutch intonation. This basic intonation pattern is said to be an abstract, mental category of intonation which underlies an actual pitch contour. This basic pattern is first proposed by 't Hart and Cohen in 1967 and was said to resemble a 'hat'. This pattern is composed of, from left to right, an initial gradual fall-off which is called a declination; a steep rise; an upward shifted segment of declination line, a steep fall, and a final declination line. The basic contour pattern for Dutch, therefore, looks like Fig. 3:

Fig. 3



"The position of rise and fall coincides with prominent syllables of those words that play a dominant part in the utterance." (Cohen and 't Hart, 1967, p. 184) Collier and 't Hart (1975) note that the intonation can only be superimposed by rule after the dominant words are known. The system of intonation cannot be expected to predict

where 'dominant' words will occur. Similarly, the present system cannot be expected to predict where the stressed syllables or 'boosted' elements will occur. The intonation system can only be applied after such decisions have been taken.

In 1975 Collier and 't Hart suggest that the basic contour can be subdivided into 'blocks' such that every intonational block contains at least one prominence-lending movement. Collier and 't Hart then continue by examining a particular type of 'block' and its function. They arrive at the conclusion that a block ending with a non-final fall marks the syntactically dominant boundary of a contour. There seems to be a fairly close correlation between the proposed tone groups in the present model and Collier and 't Hart's 'blocks', whereas the 'contour' in the Dutch system could be related to the sense group in the present system.

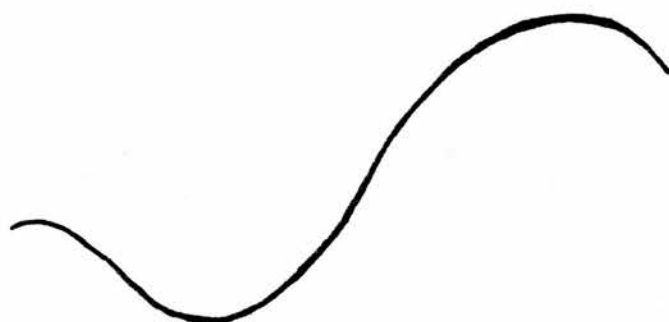
Delattre (1965) reports on his extensive investigation of German intonation. He summarises this investigation as follows:

"A complete frequency contour of German continuation or finality can be divided into five significant segments, symbolised by a bird-shaped outline seen from profile. The five segments have the following pitch shapes: a tail fall, a back depression, a neck rise, a head hump, and a beak rise, or fall or high plateau

or low plateau." (p. 158)

A basic contour configuration of the 'bird' type might look like the following:

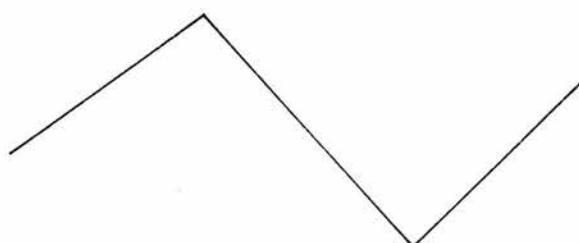
Fig. 4.



Delattre goes on to describe different contour types together with their respective functions as expressions of minor continuation, major continuation, and finality.

The above contour is very similar to the proposed contour shape for Glasgow speech, which looked like the following (cf. 7.3):

Fig. 5.



In German, the 'back' depression plus the following rise characterises the stressed syllables of an utterance, whereas in Glasgow English a similar depression preceding a rise characterises only the final stressed syllable of the tone group, other stressed syllables are realised on peaks of prominence similar to those found in Edinburgh English.

Svetorarova (1975) also works within a Hallidayan framework when she examines intonation contours in Russian. She segments the speech signal into units according to pause ("it is worth noting that pauses basically perform the function of dividing the speech chain and of expressing the degree of connection between syntagms.") In exactly the same way, the analysis outlined here proposes that the stream of speech is divided into tone groups, sense groups, and paratones, with the length of pause determining the relationship holding between these various units. After examining these units in Halliday's terms (i.e. examining 'tonics', pretonics and post-tonics), Svetorarova examines the contour as a whole and states:

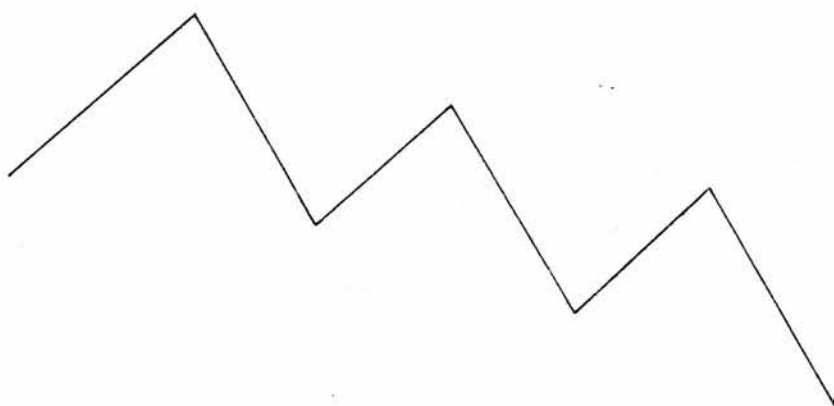
"In Russian declarative sentences there is a tendency to increase fundamental frequency on every meaningful word, except on the last and most prominent one. Against the background of relatively low and falling tones on unstressed syllables these increases in fundamental frequency, related in most cases to stressed

syllables, form a sequence of melodic peaks, the number of which is equal to the number of meaningful words in a given sentence minus one.....

Melodic rises corresponding to phonetic words may have different degrees of prominence in a sentence. In general, in declarative sentences the prominence of melodic peaks decreases from the beginning to the end of a sentence." (p. 507)

He then gives an example of this type of contour, which might be reproduced as Fig. 6:

Fig. 6



The contour pattern suggested indirectly by Svetorova looks very similar indeed to the contour pattern proposed in this thesis for various types of Scottish English, the only difference being that the baseline is more sloped in Russian than in the varieties of Scottish

English examined. A similar sloped baseline may well characterise an RP configuration.

One last model which has been worked out in similar ways to the present model, is that proposed by Thorsen (1976) for Danish intonation. The system for Danish as proposed by Thorsen is set up by examining closely the fundamental frequency configuration of short declarative sentences, interrogative sentences, and non-terminal clauses in Advanced Standard Copenhagen Danish. The model is described as follows:

"A stressed syllable and all succeeding unstressed syllables within the same simple sentence constitute a stress group, irrespective of intervening morphological and syntactical boundaries.

Sentence initial unstressed syllables constitute a separate entity and are lower than the first stressed syllable which is always relatively high.

A stressed syllable is lower than the immediately succeeding unstressed syllable (in the same stress group).

The unstressed syllable in the stress group describe a falling F_0 course. The more unstressed syllables there are the more likely the last ones are to be lower than the stressed syllable in the succeeding stress group.

The stressed syllables describe a falling contour in declarative sentences and an almost level one in interrogative sentences where the only cue distinguishing the interrogative from the declarative sentence is the intonation. The rest of the interrogative sentences and the non-terminal clauses are situated between these two extremes." (p. 85)

This model is very similar to the present model in that the stream of speech is divided up into stress groups (i.e. tone units) which are then combined into simple sentences which are realised on 'intonation contours' (p.87) (i.e. tone groups). These 'intonation contours' are then seen to vary their configuration according to their varying function, in the same way that, for example, a contour of the type X functioned as a paratone initial unit in chapter 6.

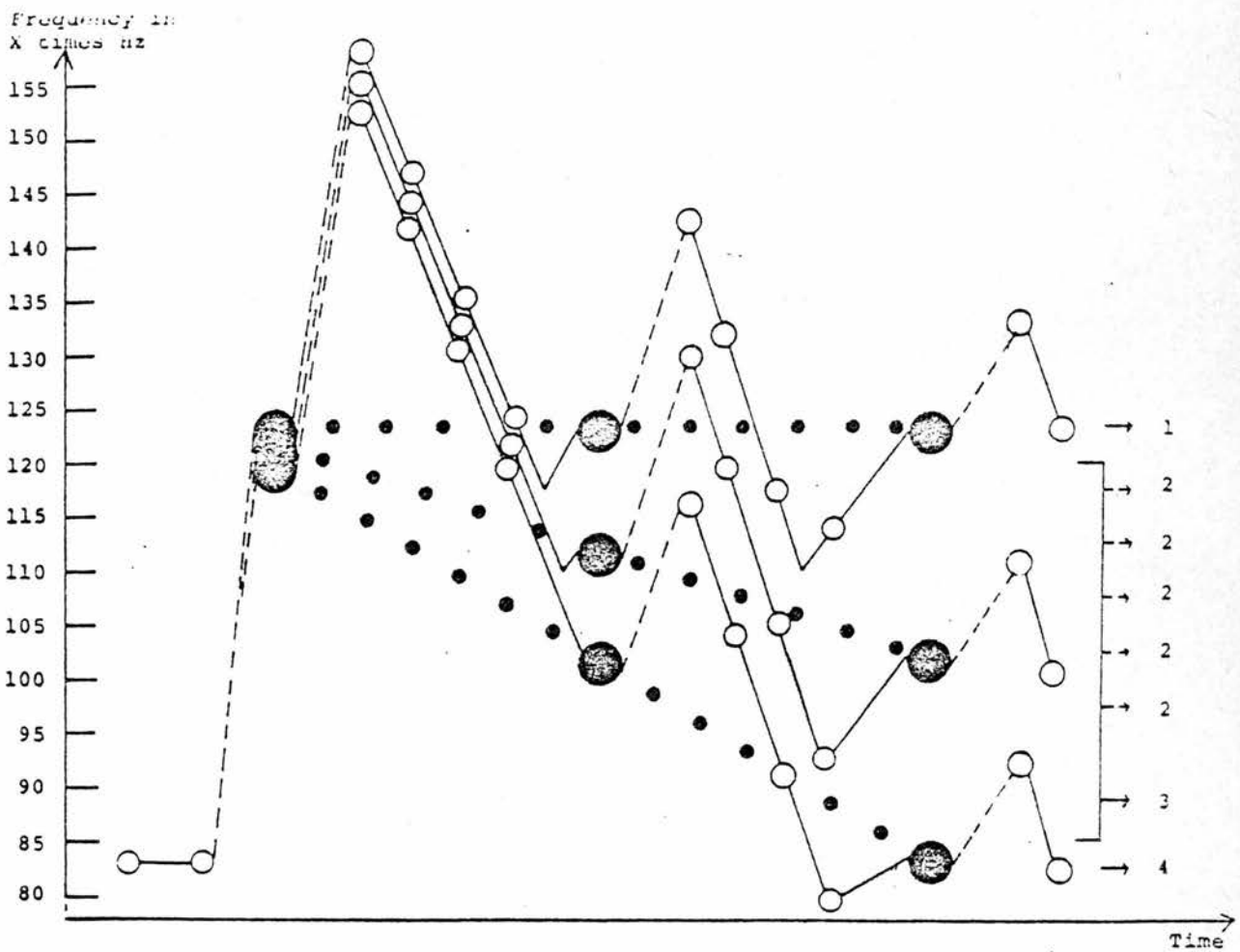
Figure 7 illustrates the various contour types described by Thorsen for Danish. This Figure represents a model for F_0 in short sentences and non-terminal clauses:

- 1) Questions where the only cue distinguishing the sentence from a statement is the intonation,
- 2) questions with word order inversion, non-terminal clauses in compound sentences,
- 3) questions with interrogative particle,

4) declarative sentences.

The stressed syllables have been marked with large filled circles; the unstressed syllables are small circles on the thin full lines. The black dots connecting the stressed syllables describe the intonation contours.

Fig. 7

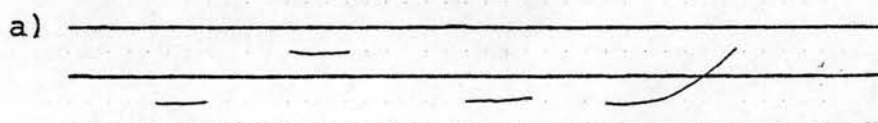


It is interesting to note that the Danish contour differs from the proposed Scottish contour in that in Danish the stressed syllables form a baseline from which the unstressed syllables depart. In the varieties of Scottish English discussed in 8.3.1, the unstressed syllables form a baseline from which the stressed syllables depart.

8.7 Conclusion

Section 8.6 has demonstrated that there is ample support in the literature for a contour system analysis of intonation which is based on the opposition between stress and unstress. A contour system analysis can be applied to intonation contours of various configurations. These configurations can vary across languages, as exemplified above in Dutch, German, Russian and Danish, and across accents, as exemplified by Edinburgh English and Glasgow English.

A contour system analysis can capture generalisations which cannot easily be expressed by models which do not examine contours as a whole. For example, the following contours can be said to be variations of a contour which is + high cf.:



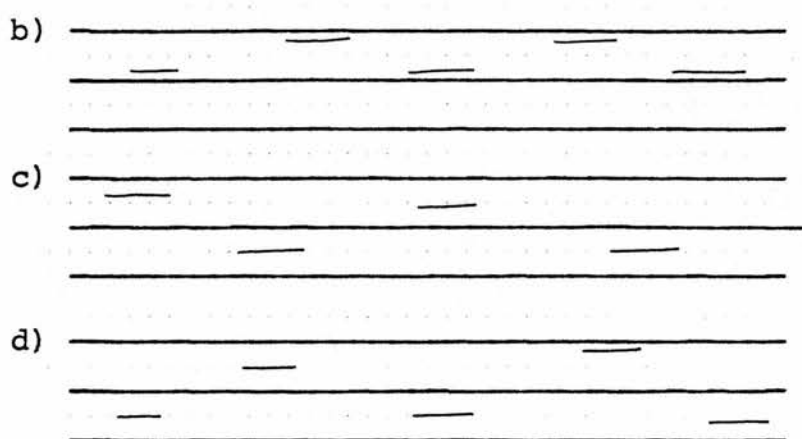


Fig. (a) contains a final rise; Fig. (b) contains a realisation of contour XY which has been raised as a whole; Fig. (c) contains an initial 'high' peak; Fig. (d) contains a final 'high' peak. All of the above apparently widely divergent patterns can be said to be realisation of a contour which contains the feature + high. These contours express various realisations of the feature + boosted height. These variations can all function in the same way in Scottish English, i.e. to express the notion 'question' (cf. Kenworthy, 1978). Note also Studdert-Kennedy (1973) who arrives at similar conclusions for Swedish contours.

A contour system analysis can be used to express generalisations at various levels of intonation analysis since there is an independent system operating at each level. Each level of analysis is defined by the distinctive oppositions contained within it as an independent system. Each system is typified by its intrinsic defin-

ing features and each system has a particular function(s). The variations in stress vs. unstress patterns can be examined at the level of the tone unit. The function of different configurations of various contour types can be examined at the level of the tone group. The relationships holding between tone groups can be examined at the level of the sense group, and the relationships holding between sense groups can be examined at the level of the paratone.

Appendix 1

Extract from conversation 31JM

+if they were a fast bricklayer + aye if they were fast, you see + the rate of wages then was + $1/8$ d an hour + an there were no overtime and eh + no piece work + you got into trouble from the union if there were piece work + and the only way they could get the work done was with a chap on a corner + this corner + and someone else on this corner + they usually gave them a haep'ny each + an hour + for to keep the line going up for to get speed up the work (mhm) that's what they did + and made their rate $1/8\frac{1}{2}$ d an the rest of the bricklayers $1/8$ d you see + until the war started in nineteen thirty-nine you got an extra haep'ny which made it $1/8\frac{1}{2}$ d (and put the others up to $1/9$ d, no?) well that's what they done + an extra haep'ny you see (well) get more work out of the boys + but as work eh got on and on the boys were saying och to hell with you and your eight and a haep'ny we want more than that + they started agitating for more money +

Bibliography

(conventional abbreviations)

Arch. Ling.	Archivum Linguisticum
Arch. Sp.	Archives of Speech
Folia Phon.	Folia Phoniatica
J.A.S.A.	Journal of the Acoustical Society of America
J.Gen.Psychol.	Journal of General Psychology
J.Phon.	Journal of Phonetics
J.Sp.H.Res.	Journal of Speech and Hear- ing Research
Lg.	Language
Lg.&Sp.	Language and Speech
Lg.Learning	Language Learning
Lings.	Linguistics
Phon.	Phonetica
Psychol.Rev.	Psychological Review
Sp.Monog.	Speech Monographs
Stud.Ling.	Studia Linguistica

Abberton E., and Fourcin, A.J. (1971) 'First application of a new laryngograph', in Med. Biol. Illustr. 21, p. 172-82.

----, (1975) 'Visual feedback and the acquisition of intonation', in (E.H.Lenneberg and E.Lenneberg, eds., Foundations of Language Development), Academic Press,

New York.

Abe, I. (1957) 'On Japanese intonation', in *Lingua* 7, p. 183-94.

Abercrombie, D., (1967) Elements of General Phonetics, Edinburgh University Press, Edinburgh.

Adams, C.M., (1969) 'A survey of Australian English intonation', in *Phon.* 20, p. 81-130.

Bolinger, D.L., (1945) 'The minimising downskip', in *Arch. Sp.* 20, p. 40-4.

-----, (1951) 'Intonation - levels vs. configurations', in *Word* 7, p. 199-210.

-----, (1958) 'Intonation and grammar', in *Lg. Learning* 8, p. 31-7.

-----, (1961) 'Contrastive accent and contrastive stress', in *Lg.* 37, p. 83-96.

-----, (1963) 'Length, vowel, juncture', in *Lings* 1, p. 5-29.

-----, (1964) 'Around the edge of language: intonation', in *Harvard Educational Review* 34, p. 282-96.

Boomer, D.S., (1965) 'Hesitation and grammatical encoding', in *Lg. & Sp.* 8, p. 148-58.

-----, & Dittman, A.T., (1962) 'Hesitation pauses and juncture pauses in speech', in *Lg. & Sp.* 5, p. 215-20.

-----, & Laver, J.D.M., (1968) 'Slips of the tongue', in *British Journal of Disorders of Communication* 3, p. 1-12.

- Bowen, J.D. (1956) 'A comparison of the intonation patterns of English and Spanish', in *Hispana* 39, p. 30-5.
- Brazil, D. (1978) 'Discourse intonation II', in *Discourse analysis monographs 2*, English Language Research, University of Birmingham.
- Brown, R.S., (1978) 'Paratones: Their reality and realisation', in SSRC Report HR 3601/1.
- Buchtal, F. (1957) 'The functional organisation of the motor unit', in *Premier Congrès International des Sciences Neurologiques*, p. 17-41.
- Chafe, W. (1976) 'Givenness, contrastiveness, definiteness, subjects, topics and point of view', in Li, C. ed., Subject and Topic, Academic Press, N.Y., p. 25-57.
- Charleston, B.M. (1960) 'Studies in the emotional and affective means of expression in modern English', in *Swiss Studies in English* 46, Berne.
- Churcher, B.G. (1935) 'A loudness scale for industrial noise measurements', in *J.A.S.A.* 6, p. 216-226.
- , King, A.J. & Davies, H. (1934) 'The minimum perceptible change of intensity of a pure tone', in *Phil. Mag.* 18, p. 927-39.
- Cohen, A. & Hart, J. (1967) 'On the anatomy of intonation', in *Lingua* 19, 177-92.
- Collier, R. (1977) 'The perception of English intonation

- by Dutch and English listeners', in IPO Annual Progress Report 12, p. 69-74, Eindhoven, Holland.
- , & 't Hart, J. (1970) 'Basic patterns in Dutch intonation', in IPO Annual Progress Report 5, p. 73-77.
- , (1972) 'Perceptual experiments on Dutch intonation', in Proc. 7th Int. Congr. Phon. Sci., Mouton, The Hague, p. 880-4.
- Cooper, W.E. (1974) 'Contingent feature analysis in speech perception', in Perception and Psychophysics 16, p. 201-4.
- , (1974) 'Perceptuo-motor adaptation to a speech feature', in Perception and Psychophysics 16, p.229-34.
- Crompton, A.S. (1978) 'Phonological Models of Intonation', in Nottingham Linguistic Circular 7, no. 1.
- Crystal, D. (1969) Prosodic Systems and Intonation in English, CUP, Cambridge.
- , (1975) The English Tone of Voice, Arnold, Bristol.
- Delattre, P., Poenack, E. and Olsen, C. (1965) 'Some characteristics of German intonation for the expression of continuation and finality', in Phon. 13.
- Denes, P. (1959) 'A preliminary investigation of certain aspects of intonation', in Lg. & Sp. 2.
- , & Milton-Williams, J. (1962) 'Further studies in intonation', in Lg. & Sp. 5, 1-14.
- Draper, M.H., Ladefoged, P. & Whitteridge, D. (1959) 'Respiratory muscles in speech', in J.Sp.H.Res. 11.

- Esser, J. (1978) 'Contrastive intonation of German and English', in *Phon.* 35, p. 41-55.
- Fairbanks, G. & Provonost, W, (1939) 'An experimental study of the pitch characteristics of the voice during the expression of emotion', in *Sp. Monog.* 6, p. 87-104.
- Flanagan, J.L. & Saslow, M.G. (1958) 'Pitch discrimination for synthetic vowels', in *J.A.S.A.* 30, p. 435-42.
- Fletcher, H. (1934) 'Loudness, pitch and timbre of musical tones and their relation to the intensity, the frequency, and the overtone structure', in *J.A.S.A.* 6, p. 59-69.
- , & Munsen, W.A. (1933) 'loudness: its definition, measurement and calculation', in *J.A.S.A.* 5, p. 82-108.
- Fourcin, A.J. (1974) 'Regularity of vocal fold vibration' in *Proc. 8th Int. Con. on Acoustics*, p. 225.
- , (1975) 'Laryngographic examination of vocal fold vibration', in B. Wyke, ed., Ventilatory and Phonatory Control Systems, O.U.P., London.
- , & Abberton, E. (1971) 'First application of a new laryngograph', in *Med. & Biol.* 11.- 21, p.172-82.
- , (1976) 'The laryngograph and voiscopie in speech therapy', in *Proc. 6th Int. Con. Logopedics and Phoniatics*, ed. Locbell, (Karger), p.116-22.

Fry, D.B. (1955) 'Duration and intensity as physical correlates of Linguistic stress' in J.A.S.A. 27, p. 765-8.

-----, (1958) 'Experiments in the perception of stress', in Lg. & Sp. 1, p. 126-52.

-----, (1965) 'The dependence of stress judgments on vowel formant structure', in Proc. Phon. Sci. 5, p. 306-11.

Garding, E. & Abramson, A.S. (1965) 'A study of the perception of some American English intonation contours', in Studia Linguistica 19, p. 61.

Gibson, J.J. (1968) The Senses considered as Perceptual systems, London, George Allan & Unwin Ltd., Ruskin House, Museum Street.

Gimson, A.C. (1956) 'The linguistic relevance of stress' in Zeitschrift fuer Phonetik und allgemeine Sprachwissenschaft 9, p. 143-9.

-----, (1962) An Introduction to the pronunciation of English, Arnold, London.

Goldman-Eisler, F. (1961) 'The distribution of pause durations in speech', in Lg. & Sp. 4, 232-7.

-----, (1961) 'Hesitation and information in speech' in Proc. 4th London Symposium of Information Theory, (ed. C. Cherry), Butterworths, Washington.

-----, (1964) 'Hesitation, information & levels of speech production', in Reuck & O'Connor (eds.)

Disorders of Language, London.

-----, (1967) 'Sequential temporal patterns and cognitive processes in speech', in *Lg. & Sp.* 10, p. 122-32.

Gulick, W.L. (1971) Hearing: Physiology and Psychophysics, O.U.P., London.

Hadding-Koch, K. (1961) Acoustic-Phonetic Studies in the intonation of Southern Swedish, Lund.

-----, (1965) 'On the physiological background of intonation', in *Stud. Ling.* 19, p. 55.

-----, & Studdert Kennedy, M. (1963) 'A study of semantic and psychophysical test responses to controlled variations in fundamental frequency', in *Stud. Ling.* 17, p. 65.

-----, (1964) 'An experimental study of some intonation contours', in *Phon.* 2, p. 175-84.

-----, (1965) 'Intonation contours evaluated by American and Swedish test subjects', in *Proc. Phon. Sci.* V, p. 326-31.

Halliday, M.A.K. (1963) 'The tones of English', in *Arch. Ling.* 15, p. 1-28.

-----, (1967) 'Notes on transitivity and theme in English', in *JL* 3, p. 177-274.

-----, (1970) A Course in Spoken English: Intonation, O.U.P., London.

Harris, J.D. (1952) 'Pitch discrimination' in *J.A.S.A.* 24, p. 750-5.

- 't Hart (1966) 'Perceptual analysis of Dutch intonation features', in IPO Annual Progress Report 1, p. 44-51.
- , (1971) 'Concatenation of intonational blocks' in IPO Annual Progress Report, p. 21-4.
- , & Cohen, A. (1973) 'Intonation by rule: a perceptual quest', in J.Phon. 1, p. 309-27.
- , & Collier, R. (1975) 'Integrating different levels of intonation analysis', in J.Phon. 3, 235-55.
- Hockett, C.F. (1955) A manual of phonology, Indiana University Publications in Anthropology and Linguistics, Memoir II of International Journal of American Linguistics.
- , (1958) A course in modern Linguistics, N.Y., Mcmillan.
- Isacenko, A. & Schädlich, H.J. (1970) A Model of Standard German Intonation, Mouton, The Hague.
- Jespersen, O. (1961) A modern English grammar on historical principles, vol. VII-syntax, Allen & Unwin, London.
- Kenworthy, J. (1977) 'Intonation of Questions (i) and (ii)', in SSRC report HR3601/1.
- , (1978) 'The Intonation of Questions: one variety of Scottish English', in Lingua 44, p. 267-82.
- Kingdon, R. (1958) Groundwork of English Intonation, Longmans, London.

- , (1958) Groundwork of English stress, Longman, London.
- Klatt, D.H. (1973) 'Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception', in J.A.S.A. 53, p. 8-16.
- Lane, H.L., Catania, A.C. & Stevens, S.S. (1961) 'Voice level: autophonic scale, perceived loudness and effects of sidetone', in J.A.S.A. 33, p. 160-8.
- Ladefoged, P. (1958) 'Syllables and stress', in *Miscellanea Phonetica III*, p. 1-14, reprinted in (1973) Phonetics in Linguistics: A Book of Readings, eds. Jones, W.E. & Laver, J., p. 205-18, Longman, London.
- , (1961) 'Stress and Pitch', in *Le Maître Phonétique*, London. republished in 1973, Phonetics in Linguistics: A Book of Readings, eds. Jones, W.E. & Laver, J., Longman, London.
- , (1967) Three Areas of Experimental Phonetics, OUP, London.
- Lehiste, I. (1961) 'Some acoustic correlates of accent in Serbo-Croatian', in *Phon.* 7, p. 114-47.
- , (1970) Suprasegmentals, Cambridge, Mass.
- , (1976) 'Suprasegmental features of speech', in Contemporary Issues in Experimental Phonetics, ed. N. Lass, p. 225-38, Academic Press, N.Y.
- , & Peterson, G.E. (1961) 'Some basic considerations in the analysis of intonation', in J.A.S.A. 33,

p. 419-25.

Liberman, A.M., Delattre, P.C., Crestman, L.J., and

Cooper, F.S. (1956) 'Tempo of frequency change as a cue for distinguishing classes of speech sounds', in J. Exp. Psychol. 52, p. 127-37.

-----, Cooper, F.S., Shankweiler, D.R., and Studdert-Kennedy, M. (1967) 'Perception of the speech code', in Psychol. Rev. 6, p. 431-61.

Lieberman, P. (1961) 'Perturbations in vocal pitch', in J.A.S.A. 33, p. 597-603.

-----, (1965) 'On the acoustic basis of the perception of intonation by linguists', in Word 21, p. 40-54.

-----, (1967) Intonation, Perception and Language, MIT Research Monograph 38, MIT Press, Cambridge, Mass.

-----, & Michaels, S.B. (1962) 'Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech', in J.A.S.A. 34, p. 922-27.

Livant, W.P. (1963) 'Antagonistic functions of verbal pauses: filled and unfilled pauses in the solution of additions', in Lg. & Sp. 6, p. 1-4.

Majewski, W, & Blasdel, R. (1969) 'Influence of fundamental frequency cues on the perception of some synthetic intonation contours', in J.A.S.A. 45, p. 450.

Mickey, K. (1977) 'On 'High' and 'Low' in English', in Lg. & Sp. 21, p. 20-39.

Nooteboom, S.G. & Slis, I.H. (1969) 'Timing in Dutch stressed words', in IPO Annual Progress Report 4, p.52-8.

O'Connor, J.D. & Arnold, G.F. (1961), Intonation of colloquial English, Longman, London. (2nd edition, 1973).

Peck, C.W. (1969) 'An acoustic investigation of the intonation of American English', in Natural Language Studies, vol. 1, University of Michigan, Ann Arbor.

Peterson & Lehiste, I. (1960) 'Duration of syllable nuclei in English', in J.A.S.A. 32, p. 693-703.

Pike, K.L. (1945) The Intonation of American English, University of Michigan Press.

-----, (1965) 'On the grammar of intonation', in Proc. 5th Int. Con. Phon. Sci., Münster, 1964, p. 105-119. (S.Karger, Basel/N.Y.).

Plomp, R. (1966) Experiments in Tone Perception, Inst. for Perception RVD-TND, Soesterberg, The Netherlands.

Pollack, I. (1948) Studies in the loudness of complex sounds, Ph.D. dissertation, Harvard University.

-----, (1952) 'The information of elementary audio displays', in J.A.S.A. 24, p. 745-9.

Rees, M. & Urquhart, A.H. (1976) 'Intonation as a guide to reader's structuring of prose texts', Work in Progress 9, Dept. Linguistics, Edinburgh University.

- Rigault, A. (1962) 'Rôle de la durée de l'intensité et de la hauteur dans la perception de l'accent en Français', in Proc. IV Int. Congr. of Phon.Sci., Helsinki, 1961, The Hague, Mouton.
- Rosenblith, W.A. & Stevens, K.N. (1952) 'Pitch discrimination data from two psychophysical methods', in J.A.S.A. 23, p. 449.
- Sharf, D.J. (1964) 'Vowel durations in whispered and in normal speech', in Lg. & Sp. 7.
- , (1971) 'Perceptual parameters of consonant sounds', in Lg. & Sp. 14.
- Simon, C. (1974) 'Analysis of some acoustic features of the speech of children', in Speech & Hearing, 1974, University College London.
- Sledd, J. (1962) 'Notes on English stress', in First Texas Conference on Problems of Linguistic Analysis in English, ed. A.A.Hill, Austin University of Texas, p. 33-44.
- Steinberg, J.C. & Munsen, W.A. (1936) 'Deviations in the loudness judgments of 100 people', in J.A.S.A. 8, p. 71-80.
- Stetson, R.H. (1951) Motor Phonetics, North-Holland, Amsterdam.
- Stevens, S.S. (1935) 'The relation of pitch to intensity' in J.A.S.A. 6, p. 150-4.
- , (1936) 'A scale for the measurement of a psychological magnitude: loudness', in Psychol. Rev.

43, p. 405-16.

----- , & Davies, H. (1938) Hearing, N.Y. Wiley.

----- , & Volkman, J. (1940) 'The relation of pitch to frequency: a revised scale', in The American Journal of Psychology, vol. 53, July 1940, no.3, p. 329-53.

----- , & ----- , & Newman, E.B. (1937) 'A scale for the measurement of the psychological magnitude of pitch', in J.A.S.A. 8, p. 185-90.

Stockwell, R.P. (1962) 'On the analysis of English intonation', in 2nd. Texas Conference on Problems of Linguistic Analysis in English, p. 39-55, ed. A.A.Hill, Austin Univ. of Texas.

Studdert-Kennedy, M. & Hadding, K. (1973) 'Auditory and linguistic processes in the perception of intonation contours', in Lg. & Sp. 16, p. 293-305.

Svetozarova, N.D. (1975) 'The inner structure of intonation contours in Russian', in Auditory Analysis and Perception of Speech, eds. Fant, G. & Tatham, M.A.A. Academic Press, N.Y.

Thorsen, N. (1976) 'An acoustical investigation of Danish intonation: preliminary results', in Annual Report of Institute of Phonetics, University of Copenhagen, vol. 10, p. 85-149.

Trager, G.L. & Smith, H.L. (1951) 'An Outline of English Structure', in Studies in Linguistics, Occasional Papers, Battenburg Press, Washington.

- Uldall, E.T. (1962) 'Ambiguity: Question or statement? or "Are you asking me or telling me?"', in Proc. IVth Int. Cong. Phon.Sci., Helsinki, 1969, Mouton, The Hague, p. 779.
- , (1964) 'Discussion of meaning in intonation' from In Honor of Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday, 12 Sept. 1961, ed. D. Abercrombie, Fry, D.B., MacCarthy, P.A.D., Scott, N.C. & Trim, J.L.M., Longman, p. 271-9.
- Van den Berg, Jw. (1957) 'Subglottic pressure and vibrations of the vocal folds', in Folia Phon. 8, p. 1-24.
- , (1958) 'Myoelastic-aerodynamic theory of voice production', in J.Sp.H.Res. 1, p. 227-44.
- , (1960) 'An electrical analogue of the trachea, lungs, and tissues', in Acta Physiologica et Pharmacologica Neolandica, 1960, 9, p. 361-85.
- Van Katwijk, A.F.V. (1972) 'On the perception of stress', paper presented at the symposium on Intonology, Prague 1970, Phonetica III, p. 127-35.
- Wingfield, A. (1975) 'The intonation syntax interaction: Prosodic features in perceptual processing of sentences', in Structure & Process in Speech Perception, eds. Cohen, A., Nooteboom, S.C., Springer-Verlag, Berlin, p. 146.